

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»**

**Інститут прикладного системного аналізу
Кафедра математичних методів системного аналізу**

До захисту допущено
В. о. завідувача кафедри
_____ О.Л. Тимощук
«___» _____ 2020 р.

**Дипломна робота
на здобуття ступеня бакалавра
за освітньо-професійною програмою «Системний аналіз і управління»
спеціальності 124 «Системний аналіз»**

**на тему: «Моделі демографічних процесів та прогнозування чисельності
населення України»**

Виконав:
студент IV курсу, групи КА-63
Пихтін Олександр Костянтинович _____

Керівник:
В.о. завідувача кафедри, доцент
Тимощук О.Л. _____

Консультант з економічного розділу:
доцент, к.е.н.
Шевчук Олена Анатоліївна _____

Консультант з нормоконтролю:
Доцент, к.т.н. Коваленко А.Є _____

Рецензент: _____

Засвідчую, що у цій дипломній роботі
немає запозичень з праць інших авторів
без відповідних посилань.

Студент _____

Київ – 2020 року

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Інститут прикладного системного аналізу
Кафедра математичних методів системного аналізу
Рівень вищої освіти – перший (бакалаврський)
Спеціальність – 124 «Системний аналіз»
Освітньо-професійна програма «Системний аналіз і управління»

ЗАТВЕРДЖУЮ

В.О.Завідувача кафедри

_____ Оксана ТИМОЩУК

«__» _____ 20__ р.

ЗАВДАННЯ

на дипломну роботу студенту

Пихтіну Олександрю Костянтиновичу

1. Тема роботи «Моделі демографічних процесів та прогнозування чисельності населення України», керівник роботи Тимощук Оксана Леонідівна, В.о. завідувача кафедри, к.т.н., затверджені наказом по університету від «25» травня 2020 р. №1143 - с
2. Термін подання студентом роботи _____
3. Вихідні дані до роботи: статистичні показники демографічних процесів
4. Зміст дипломної роботи: було досліджено проблематику моделювання та оцінки прогнозів для демографічних процесів використовуючи знання про різнецеві рівняння та використано деякі типи авторегресійних моделей.
5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо)

Студент

Олександр Пихтін

Керівник

Оксана Тимощук

РЕФЕРАТ

Дипломна робота: 94 ст., 13 табл., 25 рис., 1 дод. та 13 джерел.

ДЕМОГРАФІЧНІ ПРОЦЕСИ УКРАЇНИ, ПРОГНОЗУВАННЯ,
РІЗНИЦЕВІ РІВНЯННЯ, ЧАСОВІ РЯДИ, EVIEWS.

Тема: Моделі демографічних процесів та прогнозування чисельності населення України

В даній дипломній роботі було досліджено проблематику моделювання та оцінки прогнозів для демографічних процесів використовуючи знання про різницеві рівняння та використано деякі типи авторегресійних моделей.

Об'єкт дослідження: демографічні процеси на території України

Предмет дослідження: процес аналізу структур математичних моделей, що з прийнятною точністю описують досліджувані демографічні процеси та дають оцінку прогнозу.

Мета дипломної роботи: зробити аналіз наявних тенденцій у демографічній ситуації України та змодельовати прогноз за 2020-2024 роки.

Методи дослідження: використовувалися деякі з методів аналізу часових рядів та побудовано авторегресійні моделі.

В даній дипломній роботі представлено результати короткострокового демографічного прогнозування, які були створені на основі даних чисельності населення України, народжуваності та смертності за період з 2002 по 2019 роки з офіційного веб-сайту Державної Служби Статистики України.

ABSTRACT

Bachelor thesis: 94 p., 13 tables, 25 fig. 1 add, and 13 references

**DEMOGRAPHIC PROCESSES OF UKRAINE, FORECASTING,
DIFFERENCE EQUATIONS, TIME SERIES, EVIEWS.**

Topic: Models of demographic processes and population forecasting in Ukraine

In this thesis the problem of modeling and estimation of forecasts for demographic processes was investigated using the knowledge of difference equations and some types of autoregressive models were used.

Object of research: demographic processes on the territory of Ukraine

Subject of research: the process of analysis of the structures of mathematical models that accurately describe the studied demographic processes and give an estimate of the forecast.

The purpose of the thesis: to analyze the current trends in the demographic situation of Ukraine and model the forecast for 2020-2024.

Research methods: some of the methods of time series analysis were used and autoregressive models were built.

This thesis presents the results of short-term demographic forecasting, which were created on the basis of data on the population of Ukraine, births and deaths for the period from 2002 to 2019 from the official website of the State Statistics Service of Ukraine.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ І ПОЗНАЧЕНЬ	9
ВСТУП	10
РОЗДІЛ 1	11
ОГЛЯД ПРЕДМЕТНОЇ ОБЛАСТІ	
1.1 Сучасні демографічні процеси	11
1.1.1 Демографічні процеси у світі.....	13
1.1.2 Демографічні процеси в Україні	16
1.2 Математичні моделі та методи прогнозування демографічних процесів	19
1.3 Програми для моделювання демографічних процесів.....	26
Висновки до розділу 1 і постановка задачі	28
РОЗДІЛ 2.....	30
АНАЛІЗ ТА ВИБІР МАТЕМАТИЧНИХ МОДЕЛЕЙ ДЛЯ ДЕМОГРАФІЧНОГО ПРОГНОЗУВАННЯ.....	30
2.1 Процес створення структури математичної моделі	30
2.1.1 Поняття структури моделі	30
2.1.2 Вимоги до експериментальних даних	31
2.1.3 Аналіз стаціонарності процесу	32
2.1.4 Аналіз лінійності процесу	33
2.1.5 Гомоскедастичність та гетероскедастичність процесу	35
2.2 Прогнозування динаміки процесів	37
2.3 Оцінювання та вибір кращої з моделей	41

Висновки до розділу	45
РОЗДІЛ 3	46
ПРОЦЕС ПОБУДОВИ МАТЕМАТИЧНИХ МОДЕЛЕЙ ДЛЯ ДЕМОГРАФІЧНИХ ПРОЦЕСІВ	46
3.1 Вступ	46
3.2 Знаходження вхідних даних	46
3.3 Моделювання показників демографічних процесів	53
3.3.1 Моделювання чисельності населення України	53
3.3.2 Моделювання показників народжуваності	55
3.3.3 Моделювання показників смертності	57
3.4 Побудова прогнозів	59
3.4.1 Побудова прогнозу для чисельності населення	59
3.4.2 Побудова прогнозу для народжуваності	61
3.4.3 Побудова прогнозу для смертності	628
Висновки до розділу	63
РОЗДІЛ 4	
ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ	64
4.1 Вступ	65
4.2 Постановка задачі проектування	66
4.2.1 Обґрунтування функцій програмного продукту	66
4.2.2 Варіанти реалізації основних функцій	67
4.3 Обґрунтування системи параметрів програмного продукту	68
4.3.1 Опис параметрів	68

4.3.2 Кількісна оцінка параметрів	71
4.3.3 Експертне оцінювання параметрів	72
4.3.4 Рівень якості альтернатив функцій	77
4.4 Економічний аналіз варіантів розробки програмного продукту	78
4.5 Рівень якості альтернатив функцій	83
Висновки до розділу	84
ВИСНОВКИ ЗА РОБОТОЮ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ	85
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	86
ДОДАТОК А (ілюстративний матеріал доповіді)	87

ПЕРЕЛІК СКОРОЧЕНЬ І ПОЗНАЧЕНЬ

PP – різницеве рівняння

AR(AP) – авторегресійне рівняння

ARMA(APKC) – авторегресійне рівняння з ковзним середнім

ARIMA(APIKC) – авторегресійне рівняння з інтегрованим ковзним

DW – критерій Дарбіна-Уотсона

АКФ – автокореляційна функція

ЧАКФ – часткова автокореляційна функція

КС – ковзне середнє

SSE(СКП) – сума квадратів похибок моделі

СП – середня похибка прогнозу

СПП – середня похибка в процентах

СаПП – середня абсолютна похибка у процентах

R – коефіцієнт детермінації

AIC – інформаційний критерій Акайке

Theile – коефіцієнт Тейла

ВСТУП

У сучасному світі все більш гострим постає питання народонаселення та розробки ефективних методів дослідження демографічних процесів. Стрімкий зріст населення планети за останні 100 років тільки підвищує актуальність даних питань. Не має жодного сумніву щодо необхідності досліджень народонаселення кожної окремої країни, так і більш широкого погляду в масштабах планети. Це клопітна робота, що має бути виконуватись урядами країн заради підвищення рівня життя населення, що є основою будь-якої держави. Бо тільки при розумному підході держава може забезпечити стабільне регулювання у всіх сферах життєдіяльності населення для створення можливостей процвітання народу як нації.

Використовуючи методи прогнозування та моделювання числових рядів можливо передбачити та запобігти утворенню проблем, що пов'язані з перенаселенням на окремих територіях, побачити більш широку картину на важливі показники життя нації, а саме: народжуваність та смертність, кількість утворених шлюбів, статеві-вікова структура населення та інші.

Математичні моделі, де люди є тільки цифрами, дають змогу не тільки зробити висновки щодо подальшого розвитку, а і скласти прогноз на майбутнє з регульованим терміном тривалості.

РОЗДІЛ 1

ОГЛЯД ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Сучасні демографічні процеси

За весь час існування людства суспільство мало справу з багатьма проблемами, що мають різні масштаби, і деякі з них ставали все більше і більше впливати на розвиток суспільства. Проблеми що пов'язані з екологічною ситуацією у світі є одними з першочергових для розв'язання людством, проте не піддається сумніву факт, що люди спровокували ці самі проблеми. Маємо ситуацію, у якій існує прямий зв'язок між демографічними процесами та екологічним станом планети. Можна стверджувати, що все довкілля планети є ресурсом, який активно використовується людиною. Збільшення чисельності населення тільки приводить до кількісного зростання його потреб, що трансформується у збільшення обсягів будь-якого виробництва, і це в свою чергу може призводити до надто високого рівня забруднення довкілля. Екологічні наслідки перенаселення планети є лише однією з сторін багатогранного питання, яким зараз займаються вчені по всьому світу, а саме – дослідження народонаселення планети, оскільки дане питання потребує особливої уваги та вирішення в майбутньому.

Демографія (дав.-гр. δῆμος — народ та лат. graphe — письмо, описання) – наука про закономірності відтворення населення у суспільно-історичній обумовленості цього процесу. Демографія вивчає відтворення загалом і його компоненти як масові соціальні процеси, їх кількісні взаємозв'язки зі статеві-віковою структурою населення, залежність від соціальних та економічних явищ, характер взаємодії змін населення із соціальним розвитком. Сучасна демографія є системою наук, що об'єднує

демографічну статистику, з якою протягом тривалого часу ототожнювали власне демографію, описову, математичну, економічну, історичну, медичну, військову, етнічну та ін. Узагальнює й синтезує досягнення окремих демографічних наук теоретичної демографії, що вивчає загальні закономірності відтворення населення. Демографія тісно пов'язана з соціологією, географією населення, економікою праці, тощо. Прийняте у суспільстві трактування демографії як лише статистики населення якоюсь мірою визначало розвиток цієї науки впродовж багатьох років та й могло закріпитися. Зараз загально прийнято, що предметом демографії є відтворення населення в його суспільно-історичній обумовленості, тобто вивчення закономірностей та соціальної зумовленості народжуваності, смертності, чисельності шлюбів, їх припинення, повторне відтворення подружніх пар (сімей), а також населення загалом як єдності цих процесів.[1] Найінформативнішими показниками будемо вважати загальну чисельність населення, народжуваність та смертність.

Після розкриття питання щодо демографії як науки, тепер необхідно перейти до дослідження сучасних демографічних проблем. Маємо велику різницю між двома поняттями, а саме: демографічний вибух та демографічна криза. Цими двома поняттями вчені почали оперувати ще у третій частині XX століття, коли став помітним процес зменшення кількості населення в економічно більш розвинених країнах та збільшення кількості населення у країнах, що розвиваються. Оскільки не завжди розвинута економіка в таких державах може забезпечити всіма необхідними для розвитку соціуму інструментами, такими як якісна медицина та освіта, результатом є неспроможність конвертувати стрімкий ріст населення на користь розвитку окремої країни. Неможливо не врахувати стрімкий процес урбанізації по всьому світу, міграція, як внутрішня та зовнішня. Маємо ситуацію, у якій зараз і є людство на цій стадії розвитку – різні регіони та різні території мають нерівномірне розподілене населення.

Усі перелічені вище проблеми вже є предметом для дебатів та плідної праці багатьох організації по всьому світу і поштовхом для урегулювання народоселення.

1.1.1 Демографічні процеси у світі

Абсолютно логічним для розуміння демографічної ситуації в Україні є рішення розглянути демографічну ситуацію у світі. Переглянемо дані з офіційних джерел та спробуємо дійти до першоджерел появи критичних демографічних проблем у світі.

Сьогодні, за підрахунками ООН населення планети приблизно складає 7,5 мільярдів, це можемо спостерігати на Рисунку 1.1.

World and regional data	
	Total population in millions 2017
Arab States	359
Asia and the Pacific	3,960
Eastern Europe and Central Asia	243
Latin America and the Caribbean	641
East and Southern Africa	581
West and Central Africa	424
More developed regions	1,260
Less developed regions	6,290
Least developed countries	1,002
World	7,550

Рисунок 1.1 – Дані по світу та окремих регіонах

Цікавим є той факт, що на початку XIX століття на Землі проживало менше одного мільярда людей. Це лише більше привертає увагу до наступного твердження, за останні два століття ріст численності населення був як ніколи стрімким за всю історію людства. За багато років темпи росту

населення нашої планети не були такими, як за останні 100 років. За даними, що наведені на рисунку 1.1 бачимо що в більш розвинених регіонах проживає значно менше людей ніж в менш розвинених регіонах, і це тільки більше загострює увагу на проблемі перенаселення деяких територій.

За даними ООН, у 1950 році населення світу становило близько 2,6 мільярдів. Ріст населення став результатом для перетину позначки у 5 мільярдів людей в 1987 році і до 6 мільярдів людей в 1999 році. Станом на жовтень 2011 року населення світу становило 7 мільярдів людей. За прогнозами, населення світу зросте на два мільярди людей протягом наступних 30 років, досягнувши 9,7 мільярда до 2050 року та від 2100 до 11 мільярдів людей.

Це швидке зростання чисельності населення пов'язане головним чином із збільшенням кількості людей дітородного віку, а також з такими факторами, як зростання народжуваності, посилення урбанізації та посилення міграції. Ці тенденції будуть критичними для майбутніх поколінь.

Населення світу поділено за регіонами таким чином: 61% - в Азії (4,7 млрд), 17% - Африці (1,3 млрд), 10% - Європі (750 млн), 8% - Латинській Америці та Карибському басейні (650 млн).), 5% - Північна Америка (370 млн) та Океанія (43 млн).

Країни з найбільшим населенням - Китай (1,44 млрд) та Індія (1,39 млрд.); їх населення становить 19% та 18% світового населення.[2] За прогнозами, Індія до 2027 року здолає Китай, займаючи перше місце в світі за кількістю населення. У період з 2019 по 2050 рік населення Китаю зменшиться на 31,1 мільйона, що становитиме приблизно 2,2% населення країни.

Якщо спробувати поглянути на картину росту населення планети не тільки за останні декілька десятиліть, то приблизно можливо виділити три етапи історії людства та зростання населення. Першим є етап, який мав місце

ще до початку існуючої історії, і відзначився доволі повільною швидкістю зростання кількості населення. Другим етапом є період, у якому середня тривалість життя зростала, що було прямим наслідком до розвитку медичної сфери, становлення економіки тощо. Зараз же, другий період закінчився і можна сказати що темпи зростання кількості населення падають, що приводить історію людства до третього етапу.

У своїх дослідженнях ООН, на які я активно посилаюсь, використовують таку важливу модель, як демографічний перехід. Змістом цього терміну є зміна типу відтворення населення. Виділяють чотири типи демографічного переходу, кожен з них є притаманним для деяких окремих груп країн. Процес демографічного переходу поділяють на етапи. Зрозуміло що для кожної окремої країни процес буде різним, що обумовлено різною демографічною ситуацією на початку демографічного переходу. Якщо поглянути на ситуацію загалом, то для всіх країн по всьому світу можливо відокремити 5 етапів, у яких відбувається демографічний перехід.

1. За тривалий час до швидкого зростання населення рівень народжуваності серед населення високий, але оскільки рівень смертності також високий, ми спостерігаємо відсутність або лише дуже малий приріст населення. Ця ситуація описує більшу частину нашої історії. Суспільство в усьому світі залишалися на першому етапі протягом багатьох тисячоліть, як довгострокова перспектива щодо надзвичайно повільного зростання населення.

2. На цьому етапі смертність падає, але рівень народжуваності все ще високий. З поліпшеннями у сфері медицини та створенням превентивних та запобіжних заходів, таких як вакцинація, тощо, здоров'я населення повільно починає поліпшуватися, а рівень смертності починає падати. Оскільки стан здоров'я населення вже покращився, але народжуваність

залишається такою ж високою, як і раніше, це етап переходу, на якому чисельність населення починає швидко зростати.

3. На третьому етапі маємо ситуацію, у якій внаслідок змін у суспільстві, значних змін у економіці можна спостерігати низький рівень смертності та зниження народжуваності. Стає зрозумілим для батьків що краще забезпечити менше дітей, але зробити це якісно. Це є причиною до зниження рівня народжуваності, а отже темпи зростання населення також починають знижуватися.

4. Четвертий етап характеризується низьким рівнем смертності та низьким рівнем народжуваності. Можна сказати що ці два рівні приблизно однакові.

5. П'ятий етап є етапом с низьким рівнем смертності та деяким зростанням рівня народжуваності.

Для досягнення оптимального розвитку всього людства наразі перед країнами є задача щодо досягнення останнього етапу демографічного переходу.

1.1.2 Демографічні процеси в Україні

Демографічна ситуація в Україні є найважливішою складовою економічної, військової, соціальної безпеки в країні. Сьогодні в соціальному та соціально-економічному житті України триває перехідний період, метою якого є створення незалежної держави, демократичного суспільства та ринкової економіки. Однак процеси, пов'язані з державотворенням та переходом до принципово нової моделі економічного функціонування, обов'язково пов'язані з наявністю гострих соціальних проблем, вирішення яких потребує часу та певних об'єктивних припущень. Тому пріоритетність та послідовність у здійсненні антикризових заходів, що виникають внаслідок

поточної ситуації, є надзвичайно важливими. Основою для здійснення збалансованої державної політики населення є законодавство України, яке передбачає гарантію та захист прав людини. Взагалі демографічний розвиток України неодноразово був перерваний чинниками пертурбаційного характеру (війни, епідемії, голод, масовий терор), які спричинили глибокі демографічні кризи. Концентрованою проекцією таких криз на демографічний розвиток українського соціуму стали демографічні катастрофи: періоди, коли різко скорочувалася народжуваність і підвищувалася смертність, а природний приріст населення змінювався на скорочення.

У XX ст. в Україні відбулись три демографічні катастрофи — всі у першій половині XX століття, кожна тривалістю 7—10 років:

- 1914—1923 рр.: Перша світова, революційна доба, епідемії і голод;
- 1930—1936 рр.: колективізація і Голодомор;
- 1937—1947 рр.: Великий терор 1937—1938 рр., Друга світова війна і голод 1946—1947 рр.[4]

Важливим показником соціального розвитку країни, який відображає її соціально-економічний та моральний статус, є демографічна ситуація, яка, на жаль, погіршилася останнім часом. Сумна динаміка зміни чисельності населення України ілюструється наведеним нижче графіком:

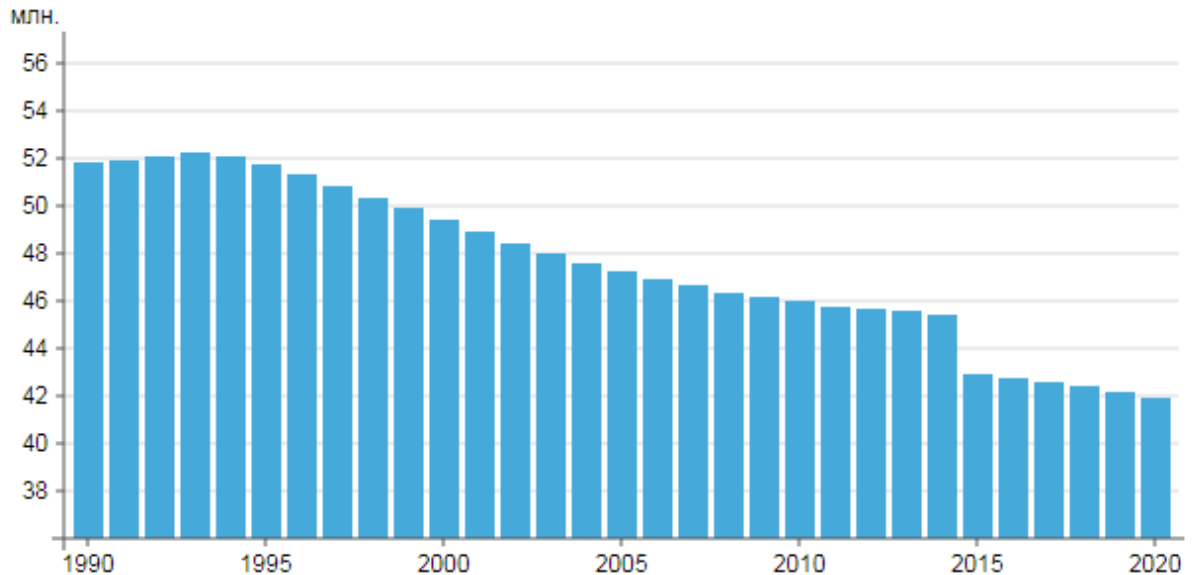


Рисунок 1.2 – Динаміка зміни чисельності населення України

Окрім позитивних явищ, для неї характерні такі негативні процеси, як: зменшення чисельності населення в країні, зниження народжуваності, високий рівень смертності, особливо в працездатному віці, погіршення здоров'я дорослих та дітей, зниження тривалості життя та поступове старіння. Призводять до різкого зростання захворюваності, необхідно назвати зміни в соціально-економічних відносинах, які впливають на зміну поведінки людей, їх ставлення до власного здоров'я. У цьому контексті особливу стурбованість викликає здоров'я дітей та дорослих працездатного віку. Спостерігається тенденція до погіршення репродуктивного здоров'я, високого рівня травматизму, професійної захворюваності та інвалідності населення.[5]

За даними Інституту демографії та соціальних досліджень ім. М.В. Птухи чисельність населення України, може скоротитися більш ніж на 12 млн. чоловік і взагалі явище депопуляції набуло окремої уваги, оскільки вже не є новиною для України. Вдалим державним регулюванням необхідно подолати тенденцію скорочення населення.

Для подолання будь-якої складної ситуації спочатку необхідно розібратись у ній, виділити причини та проблеми, намагатися глибше проаналізувати дані та зробити висновки.

Без сумніву, враховуючи тернистий шлях нашої країни у скрутні та важкі часи незалежності, внаслідок соціально-економічних перетворень держави демографічний стан країни погіршився.

Економічна ситуація у кінці останнього століття зробила значний внесок у розвиток демографічної ситуації в Україні. Внаслідок скрутного положення населення загальною ситуацією стало явище зниження рівня народжуваності. Це продовжувалось до початку нового століття. Також необхідно приймати до уваги явище міграції населення. Велика кількість працівників за кордоном не лише заробляють гроші, але й шукають способи залишитися там назавжди. Якщо мати за мету стабілізацію демографічної ситуації, то існує багато чинників, що впливають на можливість стабільного демографічного зростання, можна сказати з певністю що досягнення цієї мети є викликом. Демографів в першу чергу турбує довгострокова перспектива, оскільки міграція лише зменшує кількість населення працездатного віку.

Коротким висновком з вищезазначеного та на основі наявних знань у цій галузі, можна стверджувати, що основний спосіб досягти оптимального співвідношення між соціально-економічним та демографічним розвитком держави є не регулювання рівня народжуваності, а продовження повної економічної та соціальної активності літнього населення. Збільшення середньої тривалості життя повинно супроводжуватись підтримкою здоров'я та працездатності людини. У 2020 році потрібно більше, ніж просто прагнути покращити здоров'я людей похилого та старшого віку, а зміцнити їх розвиток таким чином, щоб забезпечити піклування як про ментальне здоров'я людини, так і про здоров'я її тіла. У цьому контексті це питання є надзвичайно важливим - розвиток адекватного відношення людей до власного здоров'я.

1.2 Основні математичні моделі та методи прогнозування демографічних процесів

В основі сучасного демографічного прогнозування лежить розробка сценаріїв для припущення можливих змін демографічних показників та прогнозування тенденцій демографічної ситуації з орієнтацією на вирішення соціально-економічних та соціальних проблем. Прогнозування зміни чисельності населення та розподілу населення є одним із важливих завдань демографії як науки. Предметом демографічного прогнозування є події та процеси відтворення населення. Предмет розрізняє прогнози майбутнього населення, зміни демографічної структури та природний та механічний рух населення. По суті, використовується науковий підхід у дослідженні та вивченні основних чинників, які є найбільш впливовими на демографічну ситуацію. Серед них – середня тривалість життя, кількість заключених шлюбів та розлучень, кількість населення та його статеві-вікова структура, зовнішня та внутрішня міграція, народжуваність та смертність, приріст населення.

Без демографічного прогнозу неможливе наукове планування показників соціально-економічному розвитку країни та її регіонів, та передбачення можливих геополітичних процесів. Висновками, які роблять демографи після проведення демографічних прогнозів користуються як державні та урядові установи, так і великі корпорації та локальні компанії. Оскільки необхідно завчасно створювати плани по виробництву товарів та послуг, розвитку інфраструктури та житлового будівництва, планування у сферах освіти та медицини та інших важливих сферах життя.

Загальноприйняті підходи, що використовуються по всьому світу експертами-демографами, є похідними від трьох основних схем. У першій

передбачається наявність існування загального закону розвитку населення, друга використовує аналіз деяких статистичних закономірностей, третя ґрунтується на вірогідній інтерпретації демографічних процесів.[4]

На сьогоднішній день задля отримання більш точних результатів в процесі демографічного прогнозування використовується об'єднання власне прогнозу та елементів моделювання. Дана технологія має назву потенційного аналізу і має два головних підходи: регіонально-топологічний та соціально-інтегративний. Соціально-інтегративний підхід має на меті зв'язати аналіз та прогноз населення, його трудового потенціалу з соціально-економічними процесами, які відбуваються і повинні розглядатись не тільки як передумови демографічних процесів, але й як їх наслідки. Регіонально-топологічний підхід розглядає необхідність подолання жорсткої прив'язки демографічних процесів до територіальної диференціації, яка традиційно вивчала основні прогнозні показники тільки в рамках адміністративно-територіального устрою. Таким чином, мета потенційного аналізу – на основі комплексної оцінки об'єктивних можливостей розвитку того чи іншого процесу визначити найбільш правдоподібні сценарії його зміни у майбутньому.[4]

На даний момент існує багато можливих класифікацій для підходів до проведення демографічного прогнозування, однією з таких може стати підхід що концентрує увагу на проблематиці та конкретно меті здійснення демографічного прогнозування:

- за часом (або горизонтом), на який розраховуються показники (короткострокові, середньострокові, довгострокові);
- за ступенем деталізації (загальна перспективна чисельність населення або за статево-віковою структурою, територіальним, соціальним розподілом тощо);
- за регіональним охопленням (глобальні – всього світу, регіональні – континентів, груп країн, національні – окремих держав, територіальні – адміністративних одиниць, великих міст);

- за призначенням або метою (практичний, прогноз-пересторога, аналітичний, нормативний тощо);
- за методикою розрахунку (екстраполяція, кореляційно-регресійні моделі тощо);
- за об'єктом, стан якого передбачається, або компонентний прогноз (чисельність окремих вікових чи соціальних груп населення, народжуваність, особливості смертності, величина міграційних потоків).[4]

Особливої уваги потребують прогнози за часом, на який розраховуються показники. Короткостроковими вважаються прогнози, що охоплюють період до п'яти років включно. Необхідність у такого роду прогнозі полягає у тому, щоб складати плани для досягнення найважливіших і актуальних цілей розвитку економічної складової та культурної складової окремих територіальних одиниць. Прогнози короткострокового тиву мають відрізнятися високим рівнем точності та проробленим рівнем деталізації.

Прогнози що мають назву середньострокових зазвичай охоплюють період від двадцяти до тридцяти років. Такого роду прогнози не завжди є точними, на відміну від короткострокових, але можуть показати глибшу картину для демографів, вказавши на перспективи росту або зменшення населення країни та її окремих частин з прийнятним рівнем надійності.\

Третій тип демографічних прогнозів, може охоплювати п'ятдесят або навіть сто років і має назву довгострокового прогнозу. Зрозуміло, що результатами такого типу пронозу оперувати можливо в якості наукових передбачень.

Ступенем деталізації є фактор, що розрізняє деталізований та не деталізований проноз. Ці два типи прогнозування відрізняються тим, що на відміну від деталізованого прогнозу, де враховується розподіл на групи (наприклад вік, стать, розміщення в місті чи за його межами, тощо) і це дає

змогу побачити ширшу картину, у не деталізованих прогнозах не має такої сегментації на групи, а є просто результат очікуваної чисельності населення.

Щодо різновидів прогнозів за призначенням, то окремо необхідно зупинитися реалістичному типі демографічного прогнозування, оскільки його використовують задля підрахунку балансів таких важливих показників як: народжуваність, смертність, а також для отримання даних щодо переходу однієї вікової групи до іншої, що є звичайним процесом.

На меті реалістичного прогнозу стоїть задача показати з високою точністю зміни, що мають пройти у соціальному житті держави та розвитку її населення. Важливо зазначити, що для проведення даного типу демографічного прогнозування використовуються довірчі інтервали(зі значеннями мінімального та максимального показників) замість сталих величин як параметрів на кожний окремий момент часу. Ступінь достовірності такого прогнозу знаходиться у прямому зв'язку з ступенем правильності прийнятої гіпотез для змін у кількості народжених, та інтенсивності змін показників смертності та напрямів міграційних потоків.

Для досягнення наукових цілей використовуються аналітичні прогнози, бо по суті, вони є вивченням тенденцій та закономірностей у процесах територіального руху населення та його відтворення. Для більшої актуальності зазвичай роблять довгострокові аналітичні прогнози, оскільки вони дають змогу продемонструвати окремий вплив обраного фактора на процес демографічного відтворення або ж розвиток населення.

Оскільки вони є довгостроковими, то не завжди відрізняються високою точністю, проте можуть забезпечити державні установи необхідними висновками щодо можливих у майбутньому негативних тенденцій. А це в свою чергу повинно стати для держави покликом для вироблення змін у напрямі проведення заходів, що стосуються демографічної політики.

У нормативному прогнозуванні головним є постановка необхідних для досягнення характеристик демографічної картини в цілому в окремій державі, зосереджується увага на показниках чисельності населення та рівні народжуваності, міграції та смертності. Наступним кроком є розробка та планування політик що приведуть до досягнення зафіксованих на початку нормативних параметрів. Такого типу прогнози фіксують увагу на зв'язку між окремими важливими демографічними показниками та прийнятими державою планами у економічному та соціально-економічному житті населення. Наприклад підвищення рівня зайнятості та добробуту населення напряду буде впливати на народжуваність та особливо на явище міграції.

Нормативне прогнозування є інтерполяційним розрахунком, бо спочатку обирається потрібне значення явища, що прогнозується, визначається форма залежності певного демографічного процесу, а потім обчислюються параметр незалежних змінних, що необхідні для досягнення бажаного результату.[4]

При здійсненні функціонального прогнозування результатом буде отримання необхідної інформації про народонаселення, достатньої для прийняття рішень в різних сферах діяльності як державних, так і соціальних установ.

На окремі групи можливо розділити наступні методи демографічного прогнозування, а саме:

- аналітичний метод;
- екстраполяційний метод;
- регресійні моделі;
- когорта-компонентний метод;
- метод марківських ланцюгів;

Аналітичний метод по своїй суті складається з припущення, що базуючись на минулій ситуації у демографічній картині держави можливо

правильно підібрати функцію, що буде найбільш точно її описувати у майбутньому.

При використанні екстраполяційного методу використовуються дані про зміни чисельності населення, як середньорічні, так і абсолютні, за окремий період або ж дані про зміну у середньорічному рівню росту населення. В основі розрахунків даного методу лежить використання експоненційної та лінійної функції. Якщо припустити що фактори або групи факторів, які можуть впливати на процес прогнозування використовуючи екстраполяційний метод є незмінними, то з'являється можливість підрахувати численність населення держави на будь-який необхідний термін. У даного методу існує значний недолік, який потребує уваги, а саме: екстраполяційного метод не враховує складні по своїй структурі зміни у окремих групах населення. Виходячи з цього зауваження екстраполяційного метод не користується популярністю при прогнозуванні наприклад вікових груп тощо.

Регресійні моделі мають місце для застосування, коли необхідно побудувати демографічний прогноз на окремий регіон. З теоретичної точки зору дана модель заснована на побудові багатовимірних регресійних моделей, які було отримано із висновків аналізу множинної кореляції та регресії.

Наступним до розгляду необхідно зазначити метод пересування вікових груп, також зустрічається за назвою як когорта-компонентний метод. Здебільшого має місце при спробах демографами прогнозувати чисельність населення. Вихідними даними для прогнозування повинні бути дані про чисельність населення та його статеві-вікову структуру. Суть методу полягає у відстеженні змін чисельності когорт на основі рівняння демографічного балансу, що дозволяє розрахувати сукупний вплив народжуваності, смертності та міграції на їх чисельність. Тому на початковому етапі необхідна побудова сценаріїв народжуваності, смертності

та міграції на основі гіпотез їх зміни. Наукове обґрунтування сценаріїв передбачає облік впливу соціально-економічних процесів, політичних змін, культурно-історичних передумов на природне і міграційний рух.

Методом марківських ланцюгів зручно користуватись, коли існує необхідність у врахуванні переходу населення з однієї групи в іншу, або ж коли частка населення за деякими причинами вибуває зі спостереження. Такими причинами можуть бути наприклад: переїзд з однієї територіальної одиниці на іншу, смерть тощо. Виходячи з принципу роботи даного методу прогнозування його також називають ергодичним, за назвою властивості ергодичності системи. Ергодичність системи – це така властивість системи, що дозволяє працювати системі ”забувши” свою попередню структуру. Таку назву метод має з причини специфіки його роботи, а саме: за припущення приймаються що коефіцієнти, які були знайдені на початку прогнозування залишаються сталими, а структура що сформована на основі цих коефіцієнтів вже від них не залежить, а залежить лише від створенної матриці переходу.

Оскільки процеси що впливають на демографічну ситуацію є не завжди лінійними та стаціонарними то все більше і більше уваги до себе привертають нові підходи до імовірнісного прогнозування. А саме користуються популярністю у демографів три основні підходи до імовірнісного прогнозування: експертне аргументоване судження, аналіз помилок історичних прогнозів, аналіз динамічних рядів.

З очевидних плюсів експертного аргументованого судження можливо виділити наступні. Цей підхід є загальноновживаним, а отже неодноразово перевірений на практиці, коли використовуючи вищезазначені методи та моделі можливо їх інтергрувати зі прогнозами, заснованими на експертній думці. Але має і свою слабку сторону, оскільки необхідно подолати суб'єктивізм експертів, і відрізнити справжні знання експерта від його суб'єктивної думки.

Підхід, що має назву аналізу помилок історичних прогнозів вимагає прискіпливої уваги до результатів попередніх демографічних прогнозів та їхнього аналізу задля побудови довірчих інтервалів для гіпотез нового прогнозу.

1.3 Програми для моделювання демографічних процесів

У сучасному світі неможливо не звертати увагу на складність, різноманітність та нелінійність демографічних процесів. Саме тому їх дослідження потребує уваги, а щоб результати цих досліджень ставали точнішими необхідно використовувати найкращі технологічні рішення, що тільки є доступними.

Основними готовими програмними продуктами, що використовуються як студентами так і вченими, для моделювання та дослідження процесів є :

- Matlab (пакет прикладних програм для числового аналізу);
- Statistica (пакет програм для всестороннього статистичного аналізу);
- SAS (набір статистичних програм, розроблений Інститутом SAS для управління даними, сучасної аналітики, багатофакторного аналізу);
- Eviews (статистичний пакет для Windows, який використовується в основному для орієнтованого на часові ряди економетричного аналізу)

Ці програмні продукти набули популярності, оскільки мають у своєму функціоналі достатньо широкий спектр можливостей для дослідження. Користувач може займатися перетворенням даних та створенням наочних графіків, діаграм, матриць і піктограм. Користувач матиме можливість дослідження та побудови різного роду часових рядів, аналізувати результати створених прогнозів з різними рівнями достовірності, тощо.

Оскільки такі пакети програмного забезпечення колись були написані на популярних та прогресивних у минулому мовах програмування, то і зараз під свої потреби кожен може розробити власний програмний продукт

використовуючи будь-яку з доступних на даний момент мов програмування, як C++, C#, Python, Java, R та інші. Доступні бібліотеки тільки роблять функціонал можливостей все більшим та зручнішим, оскільки не є продуктами, на які потрібно купувати ліцензійовану версію, на відміну від деяких готових програмних продуктів, які були зазначені вище.

Проте, експертами та демографами було створено і спеціальні пакети програмного забезпечення. Наприклад в США було створено ще в кінці попереднього століття комп'ютерну програму «Spectrum Policy Modeling System».

«Spectrum» - це система моделювання прогнозів, що підтримує аналіз, планування програм охорони здоров'я. Програма використовується для проектування майбутніх потреб та вивчення наслідків варіантів прийнятих політик.

Основний функціонал програми ділиться на декілька модулів, що розкривають потенціал програми. Розглянемо деякі з них:

1) «DemProj» – модуль що прогнозує населення для цілої країни чи регіону за віком і статтю, заснований на припущеннях про народжуваність, смертність та міграцію. Розрахунки можливо робити довгостроковими, до 100 років включно.

2) «AIM» – модуль для вивчення та аналізу наслідків епідемії ВІЛ, у тому числі кількість людей, які живуть з ВІЛ, людей що були інфіковані нещодавно, враховуючи розподіл за віком та статтю померлих від СНІДу, нові випадки захворювання на туберкульоз та дітей-сиріт. Підрахунки що зроблені цим модулем робляться кожні два роки.

3) «Human Resources» – модуль людських ресурсів дозволяє проводити калькуляцію зарплат, виплат, розраховувати стимули для постачальників медичних послуг.

4) «List» – модуль для проектування змін у рівні виживання кількості новонароджених дітей відповідно до змін в охопленні різних втручань в галузі охорони здоров'я дітей.

1.4 Висновки до розділу 1 і постановка задачі

У даному розділі було переглянуто та проаналізовано загальну ситуацію сучасних демографічних процесів на планеті та на території України. За останні 100 років кількість людей на планеті стрімко зростала, проте ріст не був рівномірним. В Південній Америці, Азійському та Африканських регіонах було відзначено зростання чисельності населення. Проте в деяких державах Європи зараз можливо побачити явище депопуляції, що є одним з етапів демографічного переходу.

Було проаналізовано демографічні проблеми що мають місце у сучасній Україні, та приведено деякі моделі та методи, що використовуються демографами по всьому світу для побудови демографічних прогнозів та передбачень. Було описано деякі програми, що використовуються для розрахунків демографічних прогнозів.

Постановка задачі:

1. Здійснити загальний опис методів та моделей демографічних прогнозів, проаналізувати їх.
2. Знайти актуальні статистичні дані за темою демографічної ситуації в Україні, використати їх для побудови описаних моделей та при прогнозуванні демографічних процесів.
3. Побудувати математичні моделі декількох видів та вказати на особливості, що відрізняють ці моделі від інших.
4. Зробити оцінку якості результатів демографічних прогнозів за відомими критеріями оцінки прогнозування.

5. Зробити висновки з отриманих результатів у ході виконання роботи, проаналізувати актуальність отриманих результатів та спробувати виробити поради щодо прийнятих на даний момент політик і заходів державою з метою покращити їх націленість у майбутньому.

РОЗДІЛ 2

АНАЛІЗ ТА ВИБІР МАТЕМАТИЧНИХ МОДЕЛЕЙ ДЛЯ ОПИСУ ДЕМОГРАФІЧНИХ ПРОЦЕСІВ

2.1 Процес створення структури математичної моделі

2.1.1. Поняття структури моделі

Перед кожним дослідником на першому кроці проведення будь-якого аналізу та моделювання процесу постає закономірне питання – якого типу цей процес? Саме тому вкрай необхідним є процедура формування моделі, за якою буде здійснюватись розрахунок. Щоб якнайкраще підготуватись, необхідно використати багато джерел про види процесів для переказу їхніх характеристик.

Зрозуміло що для демографічних процесів неможливо однозначно відповісти на це запитання, оскільки врахувати необхідно дуже багато факторів та груп факторів, що мають вплив на створення моделі.

Саме поняття структури моделі є достаньо комплексним, в нього вносять:

порядок моделі, розмірність отриманого моделлю вектора, наявність нелінійностей та її характер, запізнення реакції на виході об'єкта по відношенню до вхідного сигналу (лаговий ефект), тип збурень, який впливає на процес та способи його врахування.

В загальній ситуації, побудова моделі використовуючи часові ряди складається з наступних кроків:

- Виконати попередній аналіз вже існуючих даних, для того щоб пересвідчитися у відсутності пропусків та інших пошкоджень даних, визначити мету побудови моделі.

- Виконати перевірку існуючих часових рядів на можливу присутність нелінійностей.

- Обрати клас структур моделей-кандидатів, для цього необхідно: розрахувати и зробити аналіз кореляційної матриці для часових рядів залежної та незалежної змінних з метою виділення екзогенних змінних, які необхідно врахувати в моделі.

- Розрахувати автокореляційну та часткову автокореляційну функцію для залежної змінної з метою вибору порядку авторегресійної частини моделі.

- Обрати спосіб генерації структур моделей в залежності від обмежень на час прийняття рішення, кількості вхідних змінних.

- Обрати метод або методи оцінювання коефіцієнтів та оцінити їх параметри.

- Обрати критерій відбору кращої з моделей-кандидатів.

- Зробити перевірку на адекватність отриманої моделі в цілому.

2.1.2 Вимоги до експериментальних даних

Вимога неперервності та синхронності даних. Експериментальні дані повинні вимірюватись та реєструватись через однакові проміжки часу. Цю вимогу необхідно виконувати для процесів будь-якого типу – технічних, економічних, екологічних і т. ін. Порухення цієї вимоги призводить до зміни спектрального складу вимірювального сигналу (послідовності вимірів), що недопустимо, оскільки при цьому змінюється інформативність сигналу

Вибірка даних повинна бути представницькою. Це означає, що вона повинна охоплювати досить довгий період часу, щоб включити в розгляд всі режими роботи, які передбачається описати математичною моделлю. Розрізняють два основних режими роботи об'єктів (процесів): перехідний та усталений. В перехідному режимі система керування переводить процес з деякого початкового стану у заданий. Перебування процесу у заданому (номінальному) стані на протязі деякого відносно довгого проміжку часу називають усталеним режимом роботи.

Вибірка вимірювальних даних повинна бути інформативною. Частіше всього інформативність пов'язують з числом похідних, що їх містить вибірка даних. Чим більше число похідних можна отримати з вимірів, тим інформативнішою є часова вибірка.

2.1.3 Аналіз моделі на стаціонарність

Процеси з одиничними коренями відносять до класу нестационарних процесів. Оскільки такі процеси є досить характерними для виробничих технологій, економіки, фінансів, екології та інших галузей діяльності, то їм необхідно приділити значну увагу. Нестационарні процеси такого типу

особливо часто зустрічаються в перехідній економіці (практично всі процеси перехідної економіки нестационарні) та відповідній фінансовій діяльності, для яких властива висока нестационарна динаміка розвитку.

Сильна і слабка стаціонарність:

Розрізняють стаціонарність слабку і сильну. Слабку стаціонарність називають ще так:

- стаціонарність по коваріації;
- стаціонарність другого порядку;
- стаціонарність у широкому смислі.

Для процесу AR(1) який описується рівнянням

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k),$$

Де $\varepsilon(k)$ - послідовність білого шуму з нульовим середнім, умови стаціонарності процесу будуть такі:

- 1) модуль коефіцієнта $|a_1| < 1$ (корінь характеристичного рівняння);
- 2) однорідний розв'язок повинен дорівнювати нулю, тобто $A = 0$, або процес починається досить давно у минулому, що забезпечує збіжність $a_1^k \rightarrow 0$.

Умовою стаціонарності процесу AR(p) є рівність нулю його однорідного розв'язку.

Для процесів ковзного середнього та авторегресії довільного порядку: будь-який процес ковзного середнього, що має скінчений порядок, є стаціонарним.

2.1.4 Аналіз моделі на лінійність

Однією з проблем при визначенні структури моделі є встановлення факту наявності нелінійностей в досліджуваному процесі та їх типу. Для розв'язання цієї проблеми обов'язково використовують візуальний аналіз даних та формальні тести на наявність нелінійностей. Досвідченому фахівцю з моделювання візуальний аналіз дозволяє оперативно виявити наявність участків з лінійним або нелінійним трендом, в якійсь мірі наявність гетероскедастичності та значних імпульсних викидів (екстремальних значень), які можуть суттєво впливати на якість моделі.

Важливим моментом, який не можливо ігнорувати при побудові моделі є встановлення нелінійностей у досліджуваному процесі. Також необхідно визначити типи цих нелінійностей. Існує багато альтернативних методів для визначення нелінійності, проте необхідно знати про їхні особливості, тому зупинимося на загальному тесті Фішера.

При виконнанні побудови регресійної моделі можливо використати просту статистику (2.1)

$$\hat{F} = \frac{\frac{1}{k-2} \sum_{i=1}^k n_i (\bar{y}_i - \hat{y}_i)^2}{\frac{1}{n-k} \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2} \quad (2.1)$$

Умовними позначеннями у цій статистиці виступають такі показники:

k – кількість груп даних;

n – загальне число вимірів;

n_i – число вимірів, що були проведені в i -й групі;

\bar{y}_i – групове середнє;

\hat{y}_{ij} – оцінка по прямій регресії;

Фактично, дана статистика показує відношення відхилення середніх значень від прямої регресії к відхиленням значенням $y(k)$ від групових середніх, тобто:

$$\hat{F} = \frac{\text{Відхилення середніх значень від прямої регресії}}{\text{Відхилення значень } y(k) \text{ від групових середніх}}$$

Якщо статистика з $v_1 = k - 2$, $v_2 = n - k$ ступенями свободи є більшою за рівень значущості, то гіпотезу лінійності необхідно відкинути.

Важливо зазначити, що цією статистикою можливо користуватись тільки у тому випадку, коли структура моделі вже є заданною. Якщо наявні декілька структур, то це підвищує складність, оскільки в статистику входять оцінки \bar{y}_i

Наявність нелінійного детермінованого тренду у процесі можна визначити шляхом оцінювання рівняння:

$$y(k) = a_0 + c_1 k + c_2 k^2 + \dots + c_m k^m + \varepsilon(k),$$

яке представляє собою поліном порядку m відносно часу; випадковий процес, причини появи якого будуть розглянуті нижче у даному розділі. Якщо хоча б один із коефіцієнтів моделі, c_i $i = 1, \dots, m$, є статистично значущим, то гіпотеза щодо відсутності тренду відхиляється. Якщо тренд відносно швидко змінює свій напрямок руху і для нього важко знайти адекватний функціональний опис, то застосовують моделі випадкових трендів, які ґрунтуються на комбінаціях випадкових величин. Наявність нелінійності стосовно регресора можна встановити за допомогою відповідного полінома:

$$y(k) = a_0 + c_1 x(k) + c_2 x^2(k) + \dots + c_m x^m(k) + \varepsilon(k).$$

2.1.4 Гомо- та гетероскедастичність

Гетероскедастичними називають процеси із змінною в часі дисперсією, а гомоскедастичними – процеси із сталою дисперсією на відрізок часу, що розглядається при побудові моделі та оцінюванні прогнозу.

Гетероскедастичні процеси відносяться до широкого класу нестационарних процесів, який включає в себе такі процеси:

- процеси із детермінованими і стохастичними трендами;
- процеси із змінною дисперсією;
- процеси, які характеризуються змінним в часі математичним сподіванням та змінною дисперсією одночасно;
- процеси із змінною коваріацією (дисперсійно-коваріаційна нестационарність).

Гетероскедастичні процеси, тобто процеси із змінною в часі дисперсією, — поширений клас фінансово-економічних процесів, особливо в нестійкій перехідній економіці. За визначенням такі процеси належать до класу слабо нестационарних процесів, а тому опис їх динаміки має деякі особливості у порівнянні із стаціонарними.

Гомоскедастичні процеси – навпаки, процеси зі сталою дисперсією в часі.

Гетероскедастичність стає проблемою, коли значення змінних у рівнянні регресії значно відрізняються в різних спостереженнях. Якщо істинна залежність описується рівнянням регресії звичайним і зміни значень невиключених змінних, і похибки вимірювання, впливаючи на випадковий член, роблять його порівняно малим при малих y та x і порівняно більшим – при великих y та x , то економічні змінні часто сумісно змінюють свій масштаб.

Дуже часто появу проблеми гетероскедастичності можна передбачити заздалегідь, засновуючись на знанні характеру даних. У таких випадках можна застосувати відповідних заходів щодо усунення цього ефекту на етапі специфікації моделі регресії, і це дозволить зменшити, або, можливо, усунути необхідність формальної перевірки. До сьогодні для такої перевірки запропоновано велику кількість тестів (і критеріїв). Найбільш поширені тести:

- тест рангової кореляції Спірмена,
- тест Голдфелда-Квандта,
- тест Глейзера,
- тест Уайта.

Розглянемо процедуру тестування процесу на гетероскедастичність Глейзера:

Тест Глейзера дозволяє дещо більш ретельно розглянути характер гетероскедастичності. Ми знімаємо припущення про те, що σ_i пропорційно x_i , і хочемо перевірити, чи може бути більш підходяща будь-яка інша функціональна форма, наприклад:

$$\sigma_i = \alpha + \beta x_i^{\gamma}$$

Щоб використати даний метод, слід оцінити регресійну залежність у від x за допомогою звичайного МНК, а потім обчислити абсолютні величини залишків $|e_i|$ за функцією для даного значення γ . В кожному випадку нульова гіпотеза про відсутність гетероскедастичності буде відхилена, якщо оцінка β значущо відрізняється від нуля. Якщо при оцінюванні більш ніж однієї функції виходить значуща оцінка β , то орієнтиром при визначенні характеру гетероскедастичності може служити найкраща з них

2.2 Прогнозування динаміки процесів

Прогнозування розвитку подій у процесі приватної та ділової активності – це суттєва складова повсякденного життя, яка стосується кожного із нас. Навряд чи знайдеться такий напрям людської діяльності, де не потрібно робити прогнози розвитку подій, ситуацій, показників і т. ін. Для того щоб продемонструвати важливість коректного розв'язання задачі прогнозування розглянемо ієрархію процесу прийняття рішень на рівні виробничого, торговельного або іншого підприємства

З розвитком технологій тепер можливо прогнозувати наступні складові процесу:

- детермінованого тренду (поточного середнього значення, яке можна описати деякою вибраною функцією), як індикатора довгострокових змін процесу;

- випадкового (нерегулярного) тренду, як показника коротко- та середньострокових змін;

- короткострокових змін, тобто прогнозування коливань (відхилень), що накладаються на тренд;

- сезонних ефектів;

- приростів (швидкості) зміни процесу, які визначаються першими різницями;

- дисперсії або стандартного відхилення, як міри розсіювання процесу (наприклад, волатильність, яку часто використовують за міру ризику у інвестуванні або міру якості продукції на виробництві);

- якісних змінних процесів довільної природи (за допомогою нечіткої логіки, мереж Байєса і т. ін.);

- комбінацій вказаних елементів процесів.

Розглянемо детальніше процес побудови функції прогнозування на основі розв'язку різницевого рівняння:

Для прикладу оберемо рівняння АРКС(1.1):

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k) + \beta_1 \varepsilon(k-1), \quad |\alpha_1| < 1,$$

де $E(k)$ - білий шум з нульовим середнім;

$y(0) = y_0$ відома початкова умова.

Для однорідного рівняння

$$y(k) - a_1 y(k-1) = 0$$

розв'язком є Aa_1^k , де A – довільна константа.

Частковий розв'язок можна знайти за допомогою лагового оператора L у вигляді:

$$y(k) = \frac{a_0}{1-a_1} + \frac{\varepsilon(k)}{1-a_1L} + \frac{\beta_1 \varepsilon(k-1)}{1-a_1L}.$$

Використовуючи властивості лагового оператора, запишемо загальний розв'язок так:

$$y(k) = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i) + \beta_1 \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i-1) + A a_1^k.$$

Для того щоб знайти значення довільної константи скористаємось початковою умовою:

$$k=0: \quad y_0 = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon(-i) + \beta_1 \sum_{i=0}^{\infty} a_1^i \varepsilon(-i-1) + A.$$

Запишемо розв'язок із врахуванням отриманого значення довільної константи:

$$y(k) = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i) + \beta_1 \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i-1) + \left[y_0 - \frac{a_0}{1-a_1} - \sum_{i=0}^{\infty} a_1^i \varepsilon(-i) + \beta_1 \sum_{i=0}^{\infty} a_1^i \varepsilon(-i-1) \right] a_1^k.$$

Оскільки

$$\begin{aligned} \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i) &= \varepsilon(k) + a_1 \varepsilon(k-1) + a_1^2 \varepsilon(k-2) + \dots + a_1^k \varepsilon(0) + a_1^{k+1} \varepsilon(-1) + \dots, \\ -a_1^k \sum_{i=0}^{\infty} a_1^i \varepsilon(-i) &= -a_1^k \varepsilon(0) - a_1^{k+1} \varepsilon(-1) - a_1^{k+2} \varepsilon(-2) - \dots, \end{aligned}$$

То

$$y(k) = \frac{a_0}{1-a_1} + \sum_{i=0}^{k-1} a_1^i \varepsilon(k-i) + \beta_1 \sum_{i=0}^{k-1} a_1^i \varepsilon(k-i-1) + \left[y_0 - \frac{a_0}{1-a_1} \right] a_1^k.$$

Знайдемо рівняння для оцінювання прогнозу із врахуванням того, що на момент $k=0$ відоме значення збурення $E_0[\varepsilon(0)] = \varepsilon_0$, тобто необхідно знайти $E_0[y(k)]$. Таким чином, функція прогнозу приймає такий вигляд:

$$E_0[y(0+k) | y(0), \varepsilon(0)] = \frac{a_0}{1-a_1} + E_0 \left[\sum_{i=0}^{k-1} a_1^i \varepsilon(k-i) + \beta_1 \sum_{i=0}^{k-1} a_1^i \varepsilon(k-1-i) \right] + E_0 \left[\left[y_0 - \frac{a_0}{1-a_1} \right] a_1^k \right].$$

Враховуючи, що

$$\sum_{i=0}^{k-1} a_1^i \varepsilon(k-i) = \varepsilon(k) + a_1 \varepsilon(k-1) + a_1^2 \varepsilon(k-2) + \dots + a_1^{k-1} \varepsilon(-1),$$

Та

$$\beta_1 \sum_{i=0}^{k-1} a_1^i \varepsilon(k-1-i) = \beta_1 \varepsilon(k-1) + \beta_1 a_1 \varepsilon(k-2) + \beta_1 a_1^2 \varepsilon(k-3) + \dots + \beta_1 a_1^{k-1} \varepsilon(0),$$

отримаємо:

$$\hat{y}(0+k) = E_0[y(k)] = \frac{a_0}{1-a_1} + \beta_1 a_1^{k-1} \varepsilon_0 + \left[y_0 - \frac{a_0}{1-a_1} \right] a_1^k.$$

Отримане рівняння можна розглядати як функцію прогнозування на k кроків наперед на основі інформації, наявної на момент $k = 0$.

Знайдемо функцію прогнозування на s кроків вперед на основі інформації, наявної на момент k . Спочатку зробимо заміну індексів у рівнянні.

$$\hat{y}(0+s) = E_0[y(s)] = \frac{a_0}{1-a_1} + \beta_1 a_1^{s-1} \varepsilon_0 + \left[y_0 - \frac{a_0}{1-a_1} \right] a_1^s =$$

$$= \left(\frac{a_0}{1-a_1} \right) (1 - a_1^s) + \beta_1 a_1^{s-1} \varepsilon_0 + y_0 a_1^s.$$

Тепер виконаємо оновлення часового індексу для змінних у *та е* на k одиниць вперед:

$$\hat{y}(k+s) = E_k[y(k+s)] = \left(\frac{a_0}{1-a_1} \right) (1 - a_1^s) + \beta_1 a_1^{s-1} \varepsilon(k) + y(k) a_1^s.$$

Отримане рівняння представляє собою функцію прогнозування на основі відомої інформації про процес на момент k , включно.

Використовуючи наведені вище викладки, можна записати функції прогнозування для довільної кількості кроків таким чином:

$$\begin{aligned} s=1: \quad E_k[y(k+1)] &= a_0 + \beta_1 \varepsilon(k) + a_1 y(k); \\ s=2: \quad E_k[y(k+2)] &= \left(\frac{a_0}{1-a_1} \right) (1-a_1^2) + \beta_1 a_1 \varepsilon(k) + a_1^2 y(k); \\ s=3: \quad E_k[y(k+3)] &= \left(\frac{a_0}{1-a_1} \right) (1-a_1^3) + \beta_1 a_1^2 \varepsilon(k) + a_1^3 y(k); \end{aligned}$$

При цьому

$$\lim_{s \rightarrow \infty} E_k[y(k+s)] = \frac{a_0}{1-a_1}.$$

Можна досить просто перейти від моделі АРКС(1,1) до моделі АР(1), якщо покласти $\beta_1 = 0$. Для АР(1) отримаємо функцію прогнозування у вигляді:

$$E_k[y(k+s)] = \left(\frac{a_0}{1-a_1} \right) (1-a_1^s) + a_1^s y(k).$$

По аналогії можна знайти функції прогнозування для моделей іншої структури, табулювати отримані функції і використовувати їх при прогнозуванні процесів різної природи.

2.3 Оцінювання та вибір кращої з моделей

Важливим моментом процесу прогнозування є об'єктивне визначення якості отриманого прогнозу. Оскільки оцінки прогнозів – випадкові величини, то для встановлення їх якості необхідно використовувати декілька статистичних критеріїв. Рис. 2.1 ілюструє часову вісь та відрізки часу, на яких виконується оцінювання моделі і перевірка якості прогнозу.

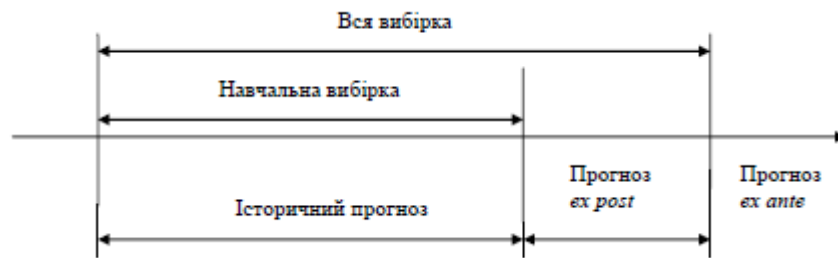


Рис. 2.1. Види прогнозування за часовим рядом

Прогноз на перевірочній частині вибірки даних в науковій літературі називають ще прогнозом *ex post*. В різних емпіричних дослідженнях рекомендують залишати для перевірки (5 – 40) % значень ряду даних. Хоча при аналізі коротких рядів для оцінювання параметрів моделі доцільно використовувати значно більшу частину ряду. Прогнозування значень поза вибіркою даних називають прогнозом *ex ante* (рис. 2.1).

Досить часто проста сума похибок прогнозів дорівнює нулю, оскільки похибки приймають різні знаки, а тому необхідно використовувати інші міри похибки. Ступінь розсіювання значень змінної навколо її середнього можна виміряти за допомогою стандартного відхилення, яке дорівнює квадратному кореню з дисперсії, тобто:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{k=1}^N [y(k) - \hat{y}(k)]^2},$$

Де $y(k)$ - фактичне значення змінної; $\hat{y}(k)$ - прогноз

Стандартне відхилення залишків – один із основних показників якості прогнозу. Це досить широко застосовувана статистична характеристика, яка виявляється корисною при аналізі поведінки процесів різної природи. Вона використовується, наприклад, як міра ризику (волатильність) при аналізі фінансових процесів, як характеристика надійності в управлінні запасами та інших прикладних задачах управління, а також як міра розсіювання значень змінних стану при аналізі систем різної природи.

Середнє абсолютне значення похибки.

Рекурсивне обчислення середнього абсолютного значення похибки (САП) ґрунтується на виразі для експоненціального середнього і має вигляд:

$$\begin{aligned} \text{САП}(k) &= \alpha \cdot |y(k) - \hat{y}(k)| + (1 - \alpha) \text{САП}(k-1) = \\ &= \alpha |e(k)| + (1 - \alpha) \text{САП}(k-1), \end{aligned}$$

$$0 < \alpha < 1, e(k) - \text{похибка прогнозу}$$

Для досить широкого класу статистичних розподілів випадкових величин значення стандартного відхилення є дещо більшим САП і строго пропорційним йому. Коефіцієнт пропорційності коливається для різних розподілів між 1,2 і 1,3.

Таким чином, можна записати, що

$$\sigma(k) \approx 1,25 \text{САП}.$$

Середній квадрат похибки і сума квадратів похибок.

Якщо середній квадрат похибок (СКП) визначається для розповсюдженого випадку однокрокового прогнозування при довжині часового ряду N , то СКП визначається за формулою:

$$\text{СКП} = \frac{1}{N} \sum_{k=1}^N [y(k) - \hat{y}(k)]^2.$$

Більшість відомих пакетів для математичного моделювання і виконання статистичних розрахунків розраховують суму квадратів похибок за виразом:

$$\text{СмКП} = \sum_{k=1}^N [y(k) - \hat{y}(k)]^2.$$

Це корисний інтегральний показник якості, який використовують для порівняльного аналізу різних методів прогнозування.

Саме СКП і СмКП – найбільш поширені критерії, які використовуються при порівняльному аналізі і виборі кращої моделі для обчислення оцінок прогнозів. Однак їх також недостатньо для поглибленого аналізу результатів прогнозування.

Середня абсолютна похибка у процентах:

Середня абсолютна похибка в процентах (САПП) – це середнє абсолютних значень похибок оцінок прогнозу в процентах відносно фактичного значення показника:

$$САПП = \frac{1}{N} \sum_{k=1}^N \frac{|y(k) - \hat{y}(k)|}{|y(k)|} \times 100\% = \frac{1}{N} \sum_{k=1}^N \frac{|e(k)|}{|y(k)|} \times 100\%,$$

або у випадку прогнозування на кроків відносно го моменту:

$$САПП = \frac{1}{s} \sum_{i=1}^s \frac{|y(k+i) - \hat{y}(k+i, k)|}{|y(k+i)|} \times 100\% = \frac{1}{s} \sum_{i=1}^s \frac{|e(k+i)|}{|y(k+i)|} \times 100\%.$$

Оскільки ця міра характеризує відносну якість прогнозу, то її використовують, в основному, для порівняння точності прогнозів різнорідних об'єктів (процесів) прогнозування.

Коефіцієнт нерівності Тейла:

Коефіцієнт нерівності Тейла U – це важливий індикатор якості моделі і прогнозу; за означенням: $0 \leq U \leq 1$. Якщо $U = 1$, то модель має практично нульові (неприйнятні) прогнозуючі властивості, що випливає із формули для обчислення U :

$$U = \frac{\sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - \hat{y}(k)]^2}}{\sqrt{\frac{1}{N} \sum_{k=1}^N y^2(k)} + \sqrt{\frac{1}{N} \sum_{k=1}^N \hat{y}^2(k)}}.$$

При прогнозовані значення співпадають з фактичними значеннями ряду – модель ідеальна. Тобто U дає можливість встановити придатність моделі (методу) в принципі для оцінювання прогнозу. Емпіричні дослідження свідчать, що високоякісним оцінкам прогнозів наближено відповідають значення $U \leq 0,005$

Коефіцієнт Тейла можна розкласти на три складові; складова, пропорційна зміщенню:

$$U^M = \frac{(\bar{y} - \hat{\bar{y}})^2}{\frac{1}{N} \sum_{i=1}^N [y(i) - \hat{y}(i)]^2};$$

складова, пропорційна дисперсії:

$$U^S = \frac{(\sigma_{\text{факт}} - \sigma_{\text{модель}})^2}{\frac{1}{N} \sum_{i=1}^N [y(i) - \hat{y}(i)]^2};$$

і складова, пропорційна коваріації:

$$U^C = \frac{2(1-\rho)(\sigma_{\text{факт}} \cdot \sigma_{\text{модель}})}{\frac{1}{N} \sum_{i=1}^N [y(i) - \hat{y}(i)]^2},$$

де ρ - коефіцієнт кореляції між залишками моделі.

Чим меншим значення приймає U^M , тим кращою є модель. Якщо $U^M = 0$ то прогнозовані значення не містять зміщення і модель адекватна процесу за цим показником.

Чим менше значення U^S , тим меншим є відхилення дисперсії виходу моделі від дисперсії ряду.

U^C - є мірою корельованості фактичного і прогнозованого за моделлю ряду.

$$U^M + U^S + U^C = 1.$$

2.4 Висновки до розділу 2

У другому розділі було розглянуто процес побудови математичної моделі, а також було акцентовано увагу на важливих етапах побудови математичної моделі.

Було показано якісні характеристики моделей, було переглянуто деякі вимоги до вхідних даних, та якщо бути точнішим – то показано як досліджувати модель на стаціонарність, лінійність, гетероскедастичність.

РОЗДІЛ 3 ПРОЦЕС ПОБУДОВИ МАТЕМАТИЧНИХ МОДЕЛЕЙ ДЛЯ ДЕМОГРАФІЧНИХ ПРОЦЕСІВ

3.1 Вступ

В третьому розділі буде більш широко розкрито питання практичної частини дипломної роботи, розпочинаючи від процесу знаходження актуальних статистичних даних, що в повній мірі зможуть надавати уявлення про стан демографічної ситуації в Україні та буде зроблено її аналіз.

3.2 Знаходження вхідних даних

Усі актуальні дані було знайдено на двох інтернет-ресурсах, а саме – з сайту Державної Служби Статистики (<http://www.ukrstat.gov.ua>).

Усі отримані дані було зведено у наступні таблиці для більшої зручності у використанні.

Таблиця 3.1 Кількість населення України за роками

Рік	Кількість населення	У містах	За містом
2002	48202500	32435400	15767100
2003	48003500	32328400	15675100
2004	47622436	32146465	15475971
2005	47280800	32009300	15271500
2006	46929500	31877500	15052000
2007	46646000	31777400	14868600
2008	46372700	31668800	14703900
2009	46143700	31587200	14556500
2010	45962900	31524800	14438100
2011	45778500	31441600	14336900
2012	45633600	31380900	14252700
2013	45553000	31378600	14174400
2014	45426200	31336600	14089600
2015	42928900	29673100	13256200
2016	42760500	29585000	13175500
2017	42584500	29482300	13102200
2018	42386403	29371000	13015400
2019	42153000	29256700	12896500

Таблиця 3.2 Кількість народжуванностей та смертей за роками

Рік	Кількість народжених	Кількість померлих
2002	390700	754900
2003	408600	765400
2004	427300	761300
2005	426100	782000
2006	460400	758100
2007	472700	762900
2008	510600	754500
2009	512500	706700
2010	497700	698200
2011	502600	664600
2012	520700	663100
2013	503700	662400
2014	465900	632300
2015	411800	594800
2016	397000	583600
2017	364000	574100
2018	335874	587665
2019	308817	581114

В якості інструментів для моделювання, було вирішено використовувати програму Eviews 7, оскільки EViews пропонує академічним дослідникам, корпораціям, державним установам та студентам доступ до потужних інструментів статистики, прогнозування та моделювання через інноваційний, простий у користуванні об'єктно-орієнтований інтерфейс.

3.3 Моделювання показників демографічних процесів

3.3.1 Чисельність населення України

Для початку побудуємо графік, щоб одразу мати більш широке представлення щодо демографічної ситуації в Україні.

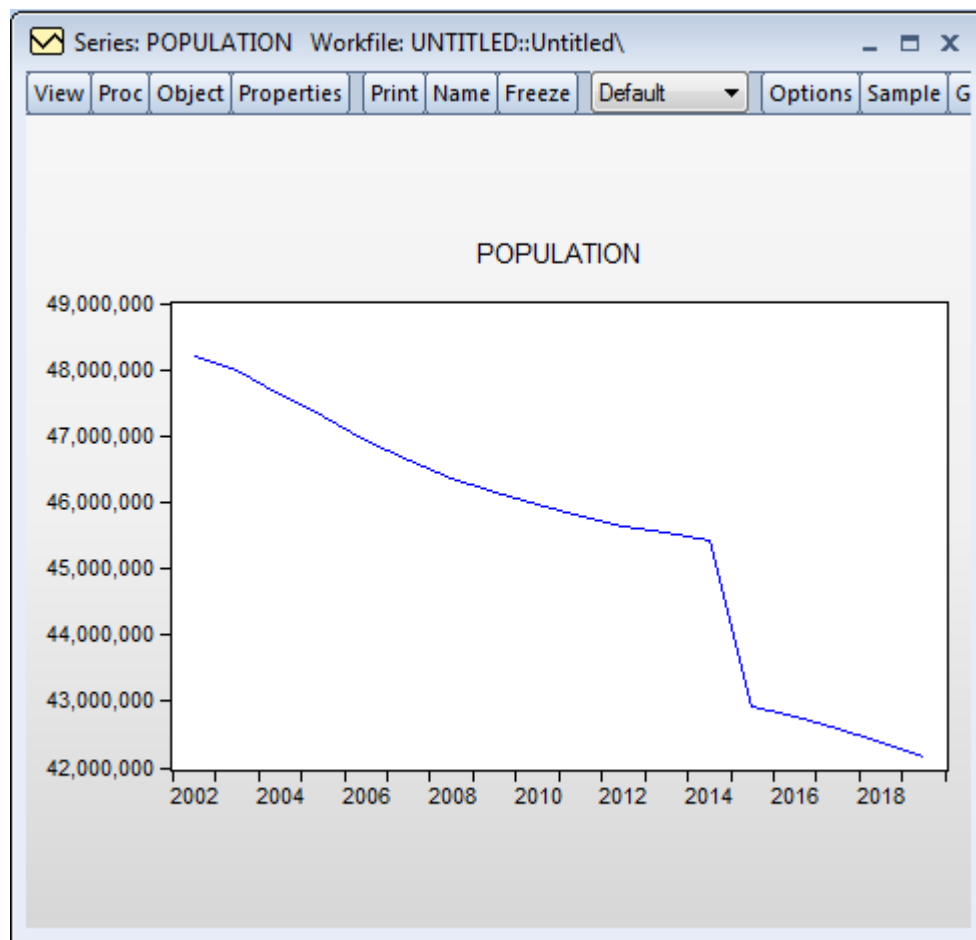


Рисунок 3.3 Численність населення України

З графіка видно, що кількість населення значно зменшилась у період 2013-2014 років, це пояснюється на сайті Державної Служби Статистики тим, що у розрахунок не приймаються тимчасово окуповані території Автономної Республіки Крим і м. Севастополя.

Спочатку необхідно побудувати моделі для фактичних значень, що надали офіційні джерела. У наступних кроках виконання роботи для зменшення розмірності та створення більш зручних обчислень буде використано логарифмування даних і результати опісля доцільно буде зрівняти один з одним.

Будуємо АКФ та ЧАКФ для ряду POPULATION

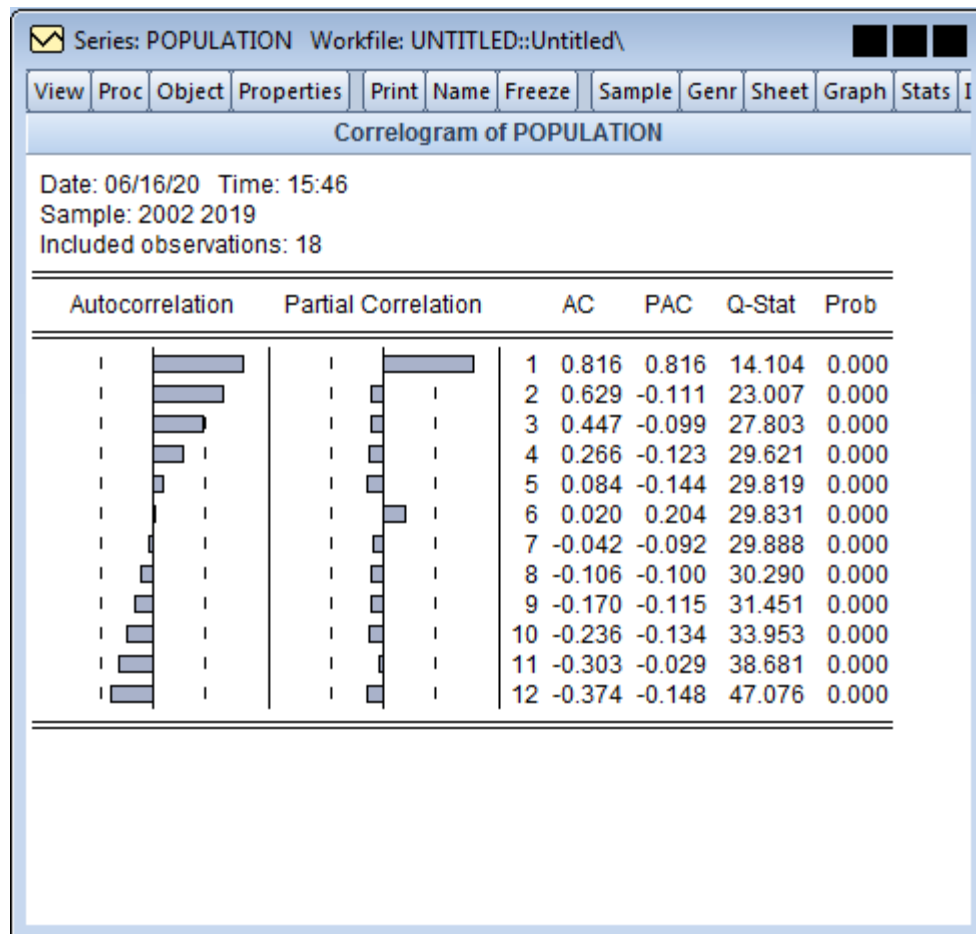


Рисунок 3.4 АКФ та ЧАКФ для ряду POPULATION

Зробивши аналіз отриманих кореляційних функцій зараз побудуємо моделі авторегресії різних порядків, а саме першого та шостого, отримаємо такі рівняння:

$$AP(1): y(k) = 45378365 + 0,967843 * y(k - 1)$$

R-squared	0.919351	Mean dependent var	45303891
Adjusted R-squared	0.913974	S.D. dependent var	1962983.
S.E. of regression	575745.3	Akaike info criterion	29.47485
Sum squared resid	4.97E+12	Schwarz criterion	29.57287
Log likelihood	-248.5362	Hannan-Quinn criter.	29.48459
F-statistic	170.9911	Durbin-Watson stat	2.206862
Prob(F-statistic)	0.000000		

Рисунок 3.5 –AP(1) (POPULATION)

R-squared	0.889887	Mean dependent var	44473659
Adjusted R-squared	0.865418	S.D. dependent var	1715327.
S.E. of regression	629275.9	Akaike info criterion	29.75485
Sum squared resid	3.56E+12	Schwarz criterion	29.87607
Log likelihood	-175.5291	Hannan-Quinn criter.	29.70996
F-statistic	36.36716	Durbin-Watson stat	2.044280
Prob(F-statistic)	0.000049		

Рисунок 3.6 –АР(6) (POPULATION)

$$AP(6): y(k) = 45436673 + 1.1030305 * y(k - 1) - 0.163211 * y(k - 6)$$

Зробивши аналіз отриманих моделей та їх чисельних характеристик вважаю за необхідне прийняти рішення щодо побудови АРКС лише беручи за основу АР(6).

Тепер щоб з'ясувати порядок ковзного середнього для АРКС моделі, що наступним кроком буде будуватися, потрібно зробити побудову АКФ та ЧАКФ для лишків, що утворилися при побудові моделі АР(6).

Autocorrelation			Partial Correlation			AC	PAC	Q-Stat	Prob	
						1	-0.062	-0.062	0.0764	0.782
						2	-0.103	-0.107	0.3061	0.858
						3	-0.074	-0.089	0.4327	0.933
						4	-0.016	-0.040	0.4394	0.979
						5	-0.023	-0.046	0.4539	0.994
						6	-0.022	-0.042	0.4683	0.998
						7	-0.022	-0.041	0.4843	1.000
						8	-0.023	-0.043	0.5031	1.000
						9	0.007	-0.014	0.5049	1.000
						10	0.019	0.002	0.5211	1.000
						11	0.060	0.052	0.7119	1.000
						12	0.002	0.009	0.7122	1.000

Рисунок 3.7 – АКФ та ЧАКФ залишків для АР(6)

Тепер можемо побудувати наступну модель:

$$АРКС(6,2): y(k) = 45347456 + 1.091256 * y(k - 1) - 0.163453 * y(k - 5) - 0.060104 * c(k - 2)$$

R-squared	0.890041	Mean dependent var	45659743
Adjusted R-squared	0.853388	S.D. dependent var	1903515.
S.E. of regression	728855.6	Akaike info criterion	30.25909
Sum squared resid	6.37E+12	Schwarz criterion	30.50415
Log likelihood	-252.2022	Hannan-Quinn criter.	30.28345
F-statistic	24.28285	Durbin-Watson stat	2.002962
Prob(F-statistic)	0.000011		

Рисунок 3.8 –АРКС(6,2) (POPULATION)

Тепер необхідно звернути увагу на той факт, що числові дані у ряді POPULATION є мільйонами, тобто кількість населення там у мільйонах, а це в свою чергу відкриває можливість для оптимізації процесу, а саме, необхідно прологарифмувати вихідні дані у POPULATION та побудувати АРКС модель використовуючи уже нові дані.

$$\begin{aligned} \text{АРКС}(6,2): \quad \log y(k) = & 18.6453 + 1.90346 * \log y(k-1) - 0.790034 * \\ & \log y(k-2) + 0.047543 * \log y(k-4) - 0.0435454 * \log y(k-5) - \\ & 0.04337 * c(k-1) - 0.034557 * c(k-2) \end{aligned}$$

R-squared	0.910287	Mean dependent var	17.63589
Adjusted R-squared	0.840509	S.D. dependent var	0.042253
S.E. of regression	0.016874	Akaike info criterion	-4.708708
Sum squared resid	0.002563	Schwarz criterion	-4.316608
Log likelihood	48.02402	Hannan-Quinn criter.	-4.669733
F-statistic	13.04563	Durbin-Watson stat	2.003402
Prob(F-statistic)	0.000465		

Рисунок 3.9 – Характеристики моделі АРКС(6,2) для новоствореного ряду

З усіх використаних і побудованих моделей тепер необхідно обрати одну, на якій буде побудовано прогноз, тому спочатку необхідно пояснити за якими критеріями буде обрано модель, що буде використана у прогнозуванні чисельності населення України.

Щоб зробити порівняльний аналіз моделей, будуть використані такі характеристики моделей:

Для порівняння моделей використовувалися наступні характеристики:

– R^2 – коефіцієнт детермінації, для кращої моделі його значення прямує до 1.

–Сума квадратів похибок моделі (SSE – sum of square error) в загальному випадку обчислюється за формулою

$$-SSE = \sum_{t=1}^N (y - \hat{y})^2,$$

де N – розмір вибірки, y – фактичні значення, а \hat{y} – значення побудовані за моделлю. Для вибору найкращої моделі необхідне мінімальне значення.

–Інформаційний критерій Акайке (AIC) обчислюється за формулою:

$$AIC = -\frac{2 \cdot l}{T} + \frac{2 \cdot k}{T}$$

де l – значення логарифмічної функції правдоподібності. Для кращої моделі $AIC \rightarrow \min$.

–Коефіцієнт Дурбина-Уотсона (DW – Durbin-Watson) показує адекватність побудованої моделі та обчислюється за формулою

$$DW = \frac{\sum_{t=2}^N (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^N \varepsilon_t^2},$$

де ε – вектор залишків (різниця між значеннями отриманими за моделлю та фактичними), при цьому $DW \in [0; 4]$. Для найкращої моделі $DW \rightarrow 2$, це означає що залишки моделі між собою не автокорелюють.

Порівняємо отримані результати, для наочності зроблено у вигляді таблиці:

Таблиця 3.3 – Порівняння моделей популяції

Модель	R^2	Sum sq _r res	DW
AP(1)	0,861416	7,58E+12	1,542153
AP(5)	0,892412	6,25E+12	1,832199
APKC(5,2)	0,891145	6,17E+12	2,004136
APKC(5,2) (log)	0.923456	-	2.003321

Маємо, що найкраще з чотирьох побудованих моделей, створених за даними POPULATION постає модель APKC(6.2), що була побудована даних POPULATION та застосуванням процесу логарифмізації цих даних.

3.3.2 Моделювання показників народжуваності

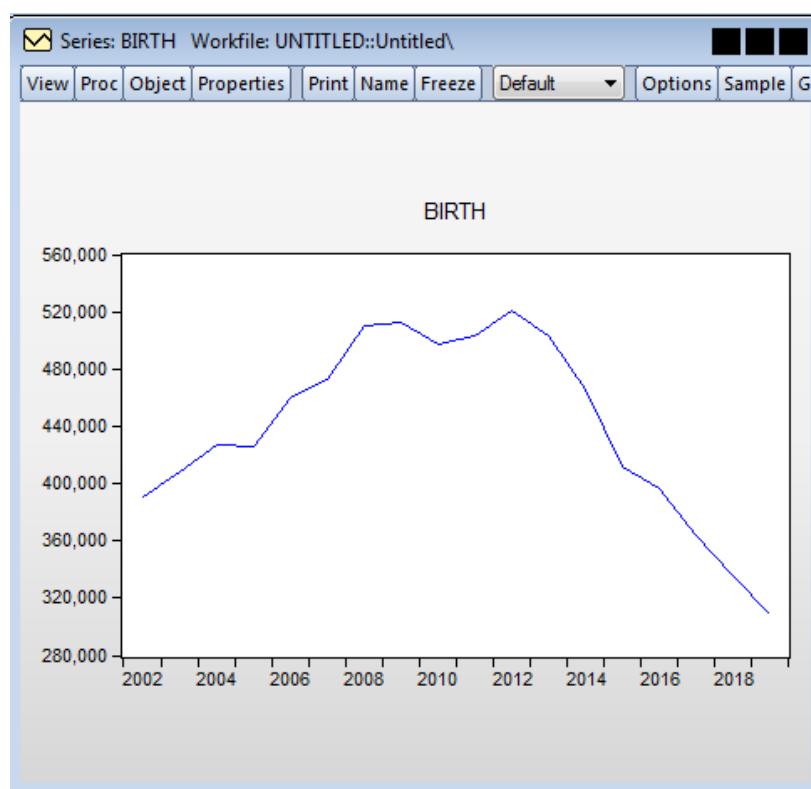


Рисунок 3.10 – Кількість народжених по роках

Зупинимося детальніше для числовому ряді BIRTH і побудуємо АКФ та ЧАКФ для цього ряду.

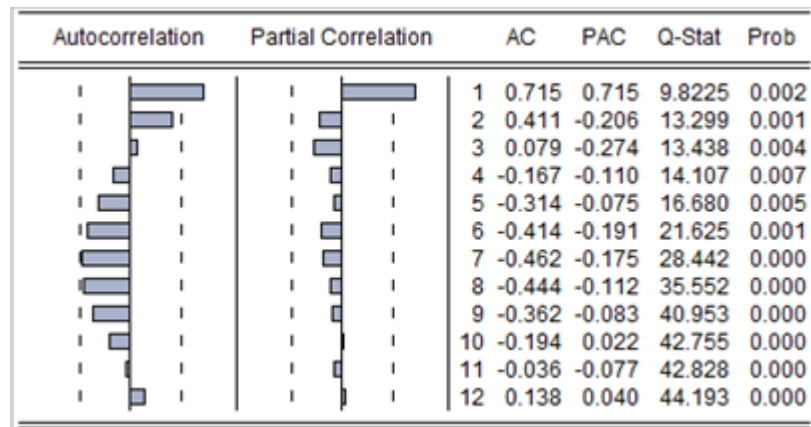


Рисунок 3.11 – АКФ та ЧАКФ (BIRTH)

Робимо висновок що необхідно будувати моделі авторегресії порядків три та сім:

$$AR(3): y(k) = 429187.6 + 1.192334 * y(k - 1) - 0.0948434 * y(k - 2) - 0.375809 * y(k - 3)$$

R-squared	0.848457	Mean dependent var	454518.8
Adjusted R-squared	0.793350	S.D. dependent var	51012.15
S.E. of regression	23189.51	Akaike info criterion	23.36290
Sum squared resid	5.92E+09	Schwarz criterion	23.60434
Log likelihood	-181.9032	Hannan-Quinn criter.	23.37527
F-statistic	15.39661	Durbin-Watson stat	1.979579
Prob(F-statistic)	0.000176		

Рисунок 3.12 – AP(3) (BIRTH)

Тепер настав час для підрахунку АКФ та ЧАКФ для залишків, що утворилися в результаті створення AP(3)

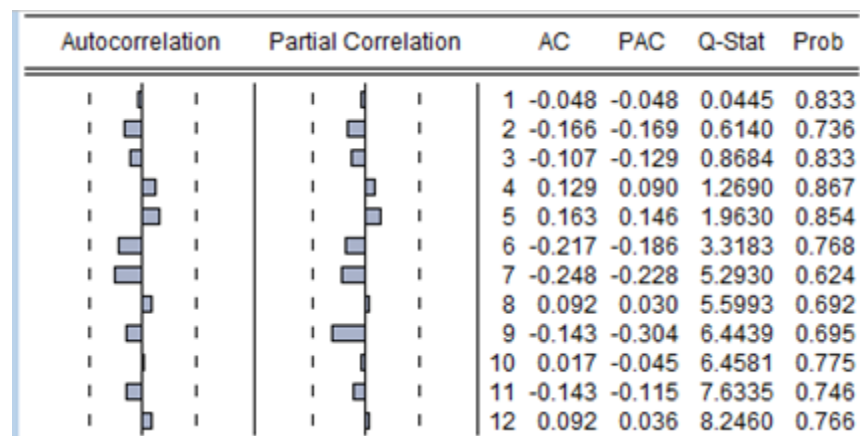


Рисунок 3.13 – АКФ та ЧАКФ залишків AP(3)

Тепер можливо побудувати наступну АРКС модель, вона матиме такий вигляд:

$$\text{АРКС}(3,3): y(k) = 458644.2 + 0.890018 * y(k-1) + 0.7867564 * y(k-2) - 0.915478 * y(k-3) - 0.488330 * c(k-2) - 0.1327850 * c(k-3)$$

R-squared	0.910790	Mean dependent var	454518.8
Adjusted R-squared	0.851317	S.D. dependent var	51012.15
S.E. of regression	19670.01	Akaike info criterion	23.29464
Sum squared resid	3.48E+09	Schwarz criterion	23.63264
Log likelihood	-179.3571	Hannan-Quinn criter.	23.31195
F-statistic	15.31427	Durbin-Watson stat	2.054667
Prob(F-statistic)	0.000291		

Рисунок 3.14 –АРКС(3,3)

Модель порядку 7 буде мати такий вигляд :

$$\text{АР}(7): y(k) = 45500243.2 + 0.7989393 * y(k-1) - 0.010383 * y(k-2) - 0.018904 * y(k-3) - 0.123322 * y(k-6) - 0.2502121 * y(k-7)$$

R-squared	0.905057	Mean dependent var	454518.8
Adjusted R-squared	0.841761	S.D. dependent var	51012.15
S.E. of regression	20292.24	Akaike info criterion	23.28833
Sum squared resid	3.71E+09	Schwarz criterion	23.62634
Log likelihood	-179.3066	Hannan-Quinn criter.	23.30564
F-statistic	14.29891	Durbin-Watson stat	1.822502
Prob(F-statistic)	0.000381		

Рисунок 3.15 –АР(7) (BIRTH)

Тепер побудуємо значення АКФ та ЧАКФ для лишків:

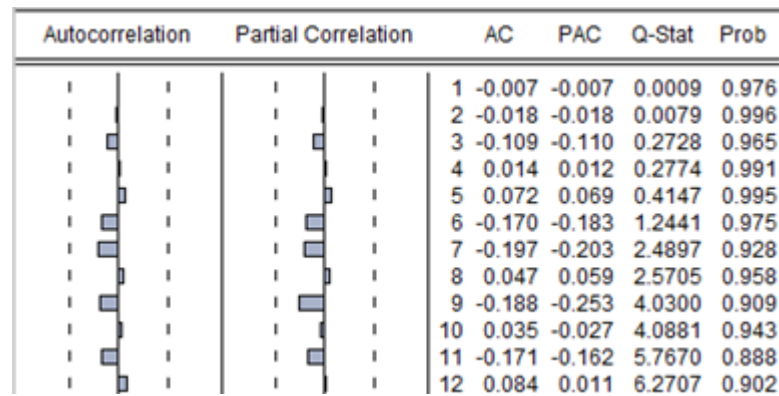


Рисунок 3.16 – АКФ та ЧАКФ залишків АР(7)

Маємо таку модель:

$$\text{АРКС}(7,7): y(k) = 456003.4 + 0.025076 * y(k-1) + 0.0085674 * y(k-2) + 0.0028764 * y(k-3) - 0.003334 * y(k-6) - 0.0134244 * y(k-7) - 0.974833 * c(k-6) - 0.0046932 * c(k-7)$$

Тепер необхідно обрати кращу з моделей, використовуючи вищеописані характеристики, для наочності всі характеристики зведемо в таблицю.

Таблиця 3.5 – Порівняння моделей народжуваності

Модель	R ²	Sum <u>sqr</u> res	DW	AIC
АР(3)	0,848457	5,92E+09	1,979579	23,36290
АР(7)	0,905057	3,71E+09	1,822502	23,28833
АРКС(3,3)	0,910790	3,48E+09	2,054667	23,29464
АРКС(7,7)	0,577394	1,65E+10	0,308124	25,19937

Найкраще себе проявила в моделюванні процесу народжуваності модель АРКС(3,3).

3.3.3 Моделювання показників смертності

Графіком продемонструємо кількість загиблих по роках.

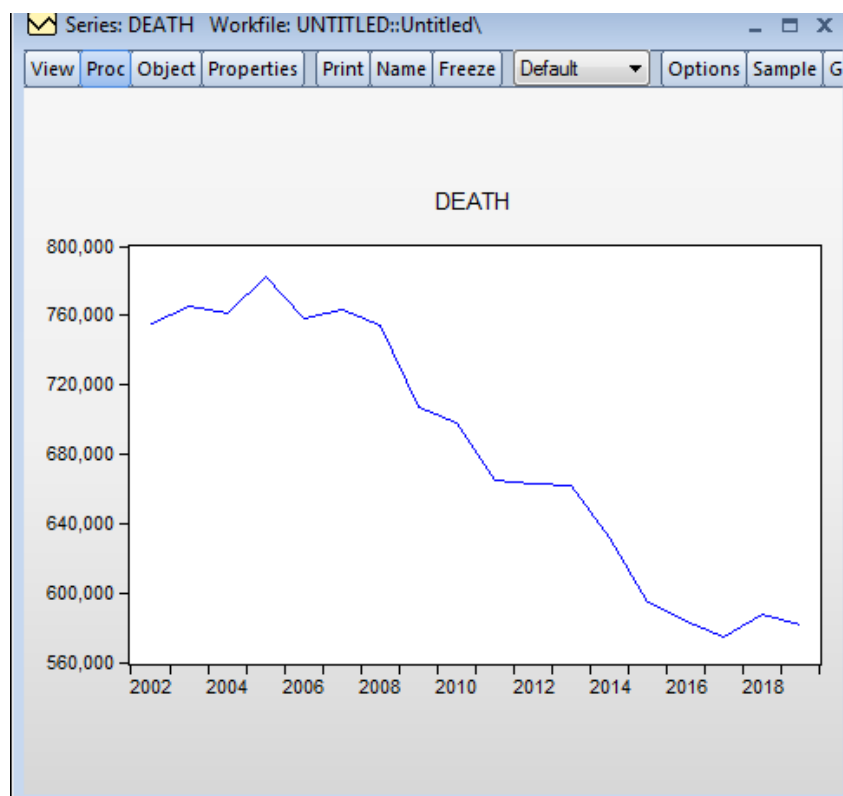


Рисунок 3.17 – Кількість загиблих по роках

Наступним логічним кроком буде побудова АКФ та ЧАКФ для ряду DEATH.

Autocorrelation		Partial Correlation		AC	PAC	Q-Stat	Prob	
				1	0.834	0.834	13.367	0.000
				2	0.649	-0.156	22.030	0.000
				3	0.447	-0.165	26.457	0.000
				4	0.257	-0.094	28.041	0.000
				5	0.123	0.045	28.440	0.000
				6	-0.050	-0.287	28.512	0.000
				7	-0.232	-0.224	30.238	0.000
				8	-0.335	0.108	34.271	0.000
				9	-0.427	-0.142	41.761	0.000
				10	-0.421	0.082	50.264	0.000
				11	-0.395	-0.053	59.254	0.000
				12	-0.374	-0.068	69.326	0.000

Рисунок 3.18 – АКФ та ЧАКФ (DEATH)

Логічним буде побудувати модель сьомого порядку, виходячи з даних Рисунок 3.18.

AP(7): $y(k) = 680675.2 + 0.778436 * y(k - 1) + 0.378389 * y(k - 2) - 0.019242 * y(k - 3) + 0.155632 * y(k - 6) - 0.644343 * y(k - 7)$

R-squared	0.954175	Mean dependent var	694931.3
Adjusted R-squared	0.923624	S.D. dependent var	71549.90
S.E. of regression	19773.62	Akaike info criterion	23.28871
Sum squared resid	3.52E+09	Schwarz criterion	23.62671
Log likelihood	-179.3096	Hannan-Quinn criter.	23.30601
F-statistic	31.23297	Durbin-Watson stat	2.021429
Prob(F-statistic)	0.000016		

Рисунок 3.19 – AP(7) (DEATH)

Тепер, наступним і логічним кроком буде вирахування АКФ та ЧАКФ для залишків з AP(7).

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.061	-0.061	0.0707	0.790
		2 -0.058	-0.062	0.1392	0.933
		3 -0.150	-0.159	0.6376	0.888
		4 -0.385	-0.424	4.1928	0.381
		5 0.318	0.266	6.8372	0.233
		6 0.000	-0.056	6.8372	0.336
		7 -0.025	-0.156	6.8575	0.444
		8 0.041	-0.029	6.9174	0.546
		9 -0.038	0.242	6.9764	0.640
		10 0.085	-0.063	7.3238	0.695
		11 0.029	-0.026	7.3739	0.768
		12 -0.073	0.035	7.7559	0.804

Рисунок 3.20 – АКФ та ЧАКФ для залишків AP(7)

В результаті отримаємо наступну модель:

АРКС(7,5): $y(k) = 673462.2 + 0.4789230 * y(k - 1) + 0.232380 * y(k - 2) + 0.4767899 * y(k - 3) - 0.304329 * y(k - 6) - 0.305658 * y(k - 7) - 0.706030 * c(k - 3) - 0.788954 * c(k - 4) + 0.285604 * c(k - 5)$

R-squared	0.978363	Mean dependent var	694931.3
Adjusted R-squared	0.945908	S.D. dependent var	71549.90
S.E. of regression	16640.83	Akaike info criterion	23.33727
Sum squared resid	1.66E+09	Schwarz criterion	23.82014
Log likelihood	-176.6982	Hannan-Quinn criter.	23.36200
F-statistic	30.14513	Durbin-Watson stat	1.744415
Prob(F-statistic)	0.000256		

Рисунок 3.21 – АРКС(7,5)

Отримані результати обох моделей тепер можливо порівняти та обрати кращу з них.

Модель	R ²	Sum sq ^{res}	DW	AIC
AP(7)	0,954175	3,52E+09	2,021429	23,28871
АРКС(7,5)	0,978363	1,66E+09	1,744415	23,33727

Оскільки коефіцієнт Дурбіна-Уотсона в моделі АРКС(7,5) не так сильно наближений до значення 2, як коефіцієнт Дурбіна-Уотсона в моделі АРКС(7) проте, модель АРКС повинна давати результати більш точні, тому важко сказати на якій з цих двох моделей необхідно зупинитись.

3.4 Побудова прогнозів

3.4.1 Побудова прогнозу для чисельності населення

Краще за всіх отриманих моделей себе проявила модель АРКС(6,2), що була побудована даних POPULATION та застосуванням процесу логарифмізації цих даних:

$$\text{АРКС}(6,2): \quad \log y(k) = 18.6453 + 1.90346 * \log y(k-1) - 0.790034 * \log y(k-2) + 0.047543 * \log y(k-4) - 0.0435454 * \log y(k-5) - 0.04337 * c(k-1) - 0.034557 * c(k-2)$$

Використовуючи кращу з отриманих моделей, а саме, модель АРКС(6,2), що була побудована даних POPULATION та застосуванням процесу логарифмізації цих даних, було отримано такі прогнозовані значення

2020 рік – 17.566802104283

2021 рік – 17.56165678979885

2022 рік – 17.55783934986241

2023 рік – 17.554388481181

2024 рік – 17.5547963289190

Тепер, щоб побачити реальні результати прогнозування, переведемо отримані результати використовуючи експоненціювання.

2020 рік – 42110413,34

2021 рік – 41836208,89

2022 рік – 41825318,02

2023 рік – 41761184,14

2024 рік – 41750323,78

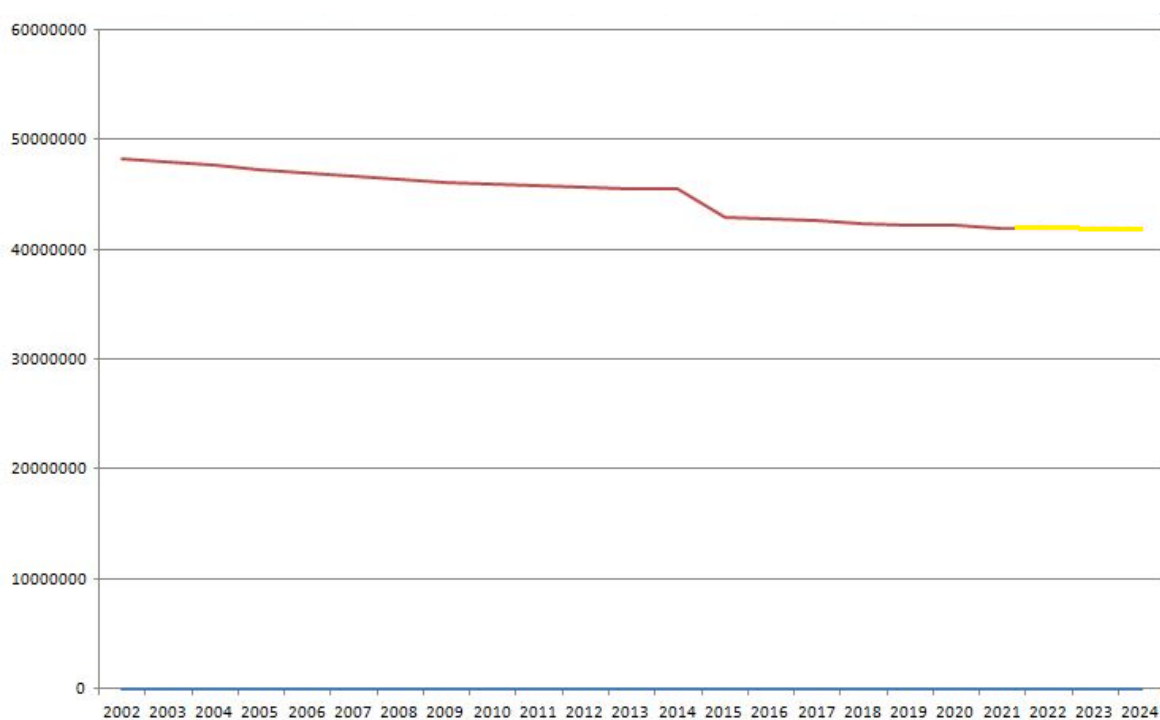


Рисунок 3.28 – Прогноз популяції

Виходячі з отриманих результатів створенного прогнозу, скоріше за все матимемо повторення минулорічних тенденцій до спадання кількості населення, проте можливо з дещо меншим темпом.

3.4.2 Побудова прогнозу для народжуваності

Спочатку оцінимо отримані прогнози:

Таблиця 3.5 Порівняння якостей прогнозів народжуваності

Модель	СерКвП	Коеф. Тейла	САПІ
AP(3)	21349.84	0.048812	7.552275
AP(7)	112336.23	0.021901	3.495452
АРКС(3,3)	25876.76	0.031455	4.951906
АРКС(7,7)	52327.12	0.058203	11.85856

Найкраще себе проявила для прогнозування процесу народжуваності модель AP(7).



Рисунок 3.29 – Графік прогнозу народжуваності

За результатами прогнозування можна стверджувати що очікується зростання рівня народжуваності.

Результати прогнозування за моделлю AP(7):

2020 рік – 390210

2021 рік – 395201

2022 рік – 400245

2023 рік – 417829

2024 рік – 429452

3.4.3 Побудова прогнозу для смертності

Модель	СерКвП	Коеф. Тейла	САПП
АР(7)	19346.41	0.015240	3.523352
АРКС(7,5)	15112.52	0.010941	1.842301

Бачимо, що логічним при прогнозуванні буде використати модель АРКС(7,5).

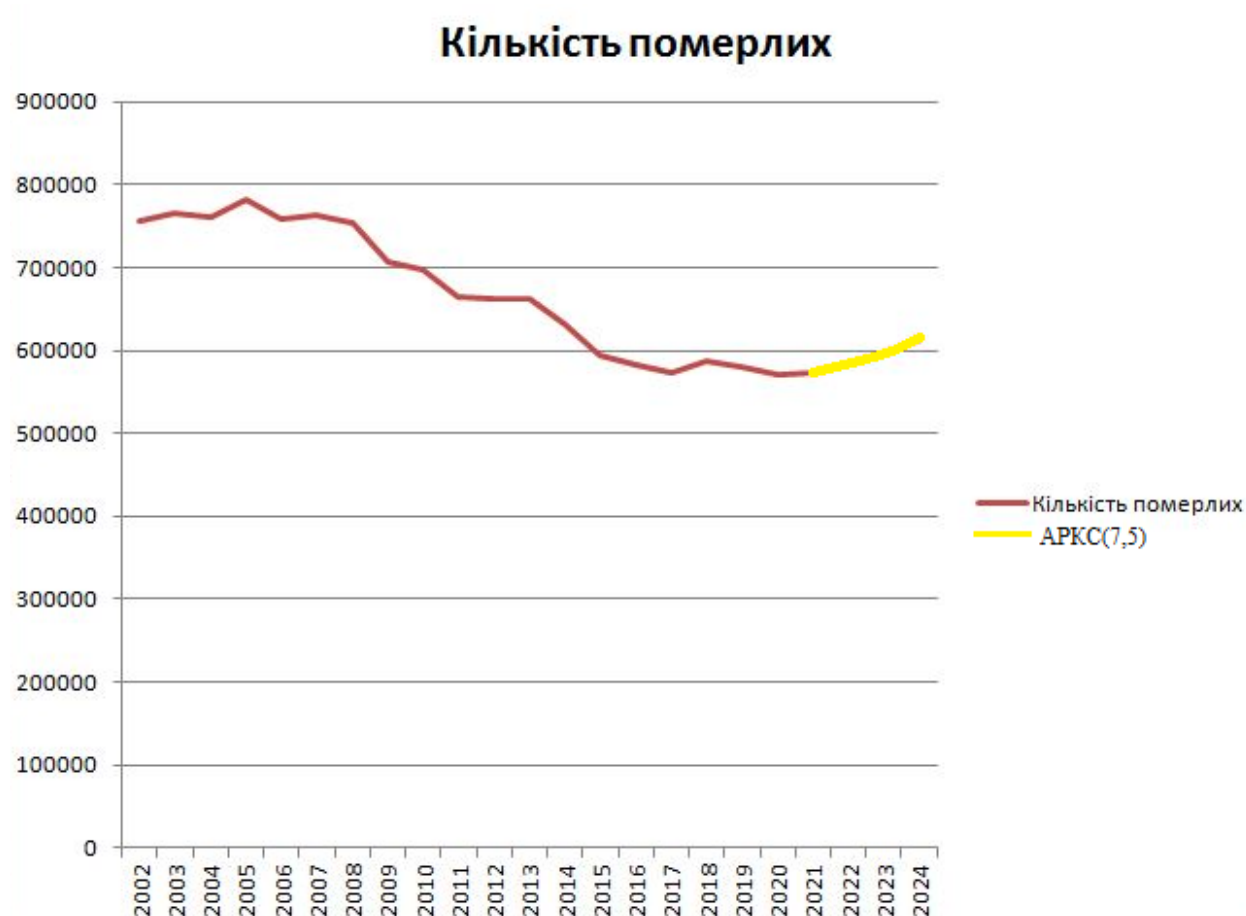


Рисунок 3.30 – Графік прогнозу смертності

Отримані дані при прогнозуванні такі:

2020 рік – 583152

2021 рік – 584669

2022 рік – 594218

2023 рік – 602161

2024 рік – 625214

Зробивши оцінку прогнозу можна побачити що результати короткострокового прогнозу на 2020-2024 рік є невтішними, відмічаються зростанням смертності.

3.5 Висновки до розділу

В третьому розділі дипломної роботи було проведено моделювання основних демографічних процесів, що можуть найкраще описати демографічну ситуацію в Україні, а також було проведено порівняння отриманих моделей за допомогою методів, що зазначені у розділі номер два.

Було створено та показано декілька різних регресійних моделей прогнозування та на їх основі зроблено короткостроковий прогноз щодо чисельності населення України, показників народжуваності та смертності.

На основі виконаної роботи було відібрано одну з моделей-кандидатів різного порядку для прогнозування. Було оцінено їх цінність, з практичної точки зору.

РОЗДІЛ 4 ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ

4.1 Вступ

Функціонально-вартісний аналіз — технологія дослідження та оцінки реальної вартості продукту, причому застосовувати її можна для будь-якої компанії з довільною структурою. Дії, що виконуються на етапах функціонально-вартісного аналізу називають функціями.

Основною метою такого аналізу є дослідження продукту на предмет максимально оптимізованої його реалізації, задля зменшення необхідної кількості часу на розробку, а також задля зниження фінансових витрат на продукт. Отже необхідно детально спланувати архітектуру продукту та усі його функціональні елементи, щоб ще перед початком розробки виключити зайві модулі та функції, що потребуватимуть зайвих витрат час та збільшать прямі і й непрямі фінансові затрати на написання програми.

Для покращення якості дослідження усіх технічних та технологічних сторін програмного продукту, усіх його модулів та функціональних частин, оцінки витрат на реалізацію всіх його функцій рекомендується залучати групу фахівців, бажано, щоб вони були експертами різних галузей і, відповідно, кожен з них відповідав за свій напрямок досліджень, в якому він є фахівцем.

Основна мета: вибір найбільш оптимального способу розробки програмного продукту для моделювання та оцінки прогнозів демографічних процесів.

4.2 Постановка задачі проектування

Необхідно оцінити можливі варіанти розробки власного програмного продукту для моделювання демографічних процесів та оцінки їх прогнозів. Основними функціональними вимогами до програми є:

- Можливість вибору типу моделі та порядок параметрів
- Зручний інтерфейс користувача
- Візуалізація вихідних даних та результатів моделювання й прогнозування
- Обчислення характеристик моделей-кандидатів для визначення їх адекватності

Також сформулюємо основні технічні вимоги до програмного продукту:

- Кросс-платформеність
- Висока швидкодія
- Невисокі вимоги до апаратного забезпечення
- Простота взаємодії з ПЗ
- Невисокі витрати на впровадження та підтримку ПЗ

4.2.1 Обґрунтування функцій програмного продукту

Головна функція F_0 – розробка програмного продукту для моделювання демографічних процесів та оцінки прогнозів з можливістю примусового вибору конкретних моделей.

Спираючись на поставлену мету, виділимо неступні функції та альтернативні варіанти їх реалізації:

F_1 – мова програмування:

- a) C++
- b) Python
- c) R

F_2 – спосіб введення вихідних даних:

- a) Зчитування з файлу
- b) Ініціалізація даних в коді програми
- c) Введення даних вручну за допомогою інтерфейсу програми

F_3 – інтерфейс користувача:

- a) Веб-інтерфейс
- b) Віконний інтерфейс
- c) Консольний інтерфейс

F_4 – виведення результатів моделювання:

- a) В окремий файл
- b) На екран з додатковою візуалізацією

4.2.2 Варіанти реалізації основних функцій

Складемо морфологічну карту варіантів реалізації базових основних функцій, що буде відображати усі можливі варіанти реалізації основних функцій та матрицю переваг та недоліків кожного варіанту функцій:

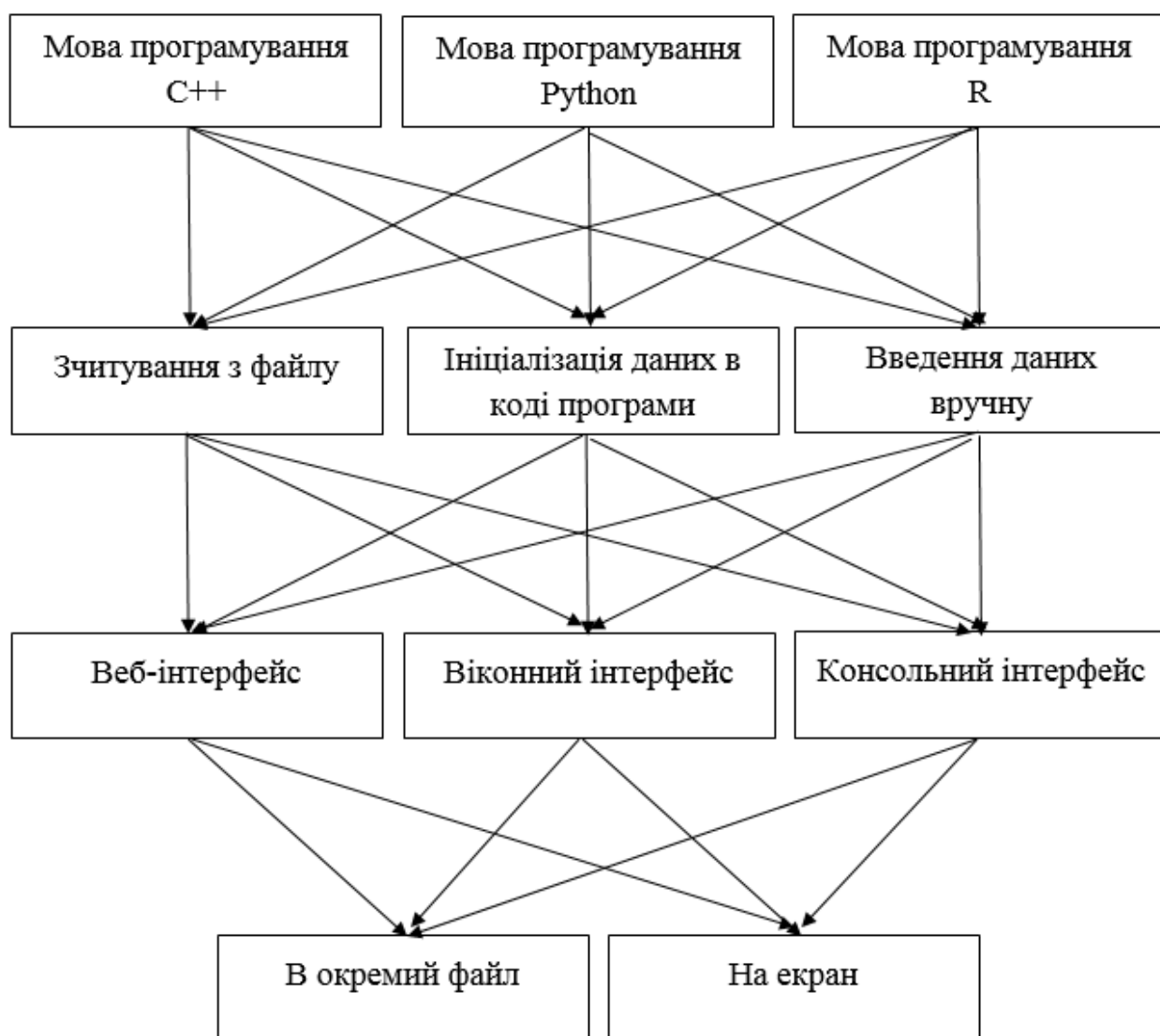


Рисунок 4.1 – Морфологічна карта

Таблиця 4.1 – Матриця переваг та недоліків

Основні функції	Варіанти реалізації	Переваги	Недоліки
F1	A	Швидкодія	Складна структура коду, багато часу на його написання
	B	Простота реалізація, зручність	Невисока швидкодія

Продовження таблиці 4.1

	С	Досить проста реалізація	Низька швидкодія
F2	А	Простота реалізації, ергономічність	Вимоги до формату файлу
	В	Легкість редагування окремих показників	Зміна усіх займає велику кількість часу
	С	Не вимагає реалізації	Високі витрати часу на введення даних
F3	А	Простий у реалізації	Необхідний веб-оглядач
	В	Не потребує додаткового ПЗ	Велика кількість часу на реалізацію
	С	Не потребує реалізації	Низька ергономічність, відсутність візуалізації
F4	А	Збереження даних	Низька ергономічність
	В	Висока ергономічність	Проблематичне збереження даних

Тепер за допомогою матриці переваг та недоліків можна обрати, які альтернативи слід відкинути для кожної з основних функцій:

Функція F1:

Оскільки час на розробку програмного продукту є обмеженим і серед необхідних досліджень є й інші задачі, то варіант а) відкидаємо, також необхідно відкинути варіант с) через низьку швидкодію.

Функція F2:

Програмний продукт повинен працювати з різними часовими рядами даних, отже альтернативи b) та c) будуть відкинуті.

Функція F3:

Для простоти комунікації користувача з програмним забезпеченням необхідно реалізувати зручний та ергономічний інтерфейс, варіант c) не відповідає цим вимогам, проте враховуючи обмеженість часу на реалізацію продукту альтернативу b) також відкидаємо.

Функція F4:

Кожну з альтернатив необхідно розглянути, оскільки їх неможливо адекватно порівняти за спільними параметрами, кожен з варіантів можна однаково швидко реалізувати та використовувати.

Отже, буде розглянено такі варіанти реалізації програмного продукту:

1. $F1b - F2a - F3a - F3a$
2. $F1b - F2a - F3a - F3b$

4.3 Обґрунтування системи параметрів програмного продукту

4.3.1 Опис параметрів

На підставі даних про основні функції, що повинен реалізувати програмний продукт, вимог до нього, визначаються основні параметри виробу, що будуть використані для розрахунку коефіцієнта технічного рівня.

Для того, щоб охарактеризувати програмний продукт, будемо використовувати наступні параметри:

- X1 – швидкодія програмного коду;
- X2 – об'єм оперативної пам'яті;
- X3 – час обробки даних;
- X4 – потенційний об'єм коду.

X1: Відображає швидкодію операцій залежно від обраної мови програмування та якості написання коду.

X2: Відображає об'єм оперативної пам'яті персонального комп'ютера, необхідний для збереження та обробки даних під час виконання програми.

X3: Відображає час, який витрачається на внутрішні операції всередині програми.

X4. Показує розмір програмного коду який необхідно створити безпосередньо розробнику.

4.3.2 Кількісна оцінка параметрів

Гірші, середні і кращі значення параметрів вибираються на основі вимог замовника й умов, що характеризують експлуатацію програмного продукту, складемо таблицю, що їх відображає:

Таблиця 4.2 – Основні параметри

Назва параметра	Умовне позначення	Одиниці виміру	Значення параметра		
			гірші	середні	кращі
Швидкодія програмного коду	X1	Оп/мс	3500 0	50000	7000 0

Об'єм оперативної пам'яті	X2	Мб	40	25	15
Час обробки даних	X3	мс	1500	800	400
Потенційний об'єм коду	X4	Кількість рядків коду	800	600	400

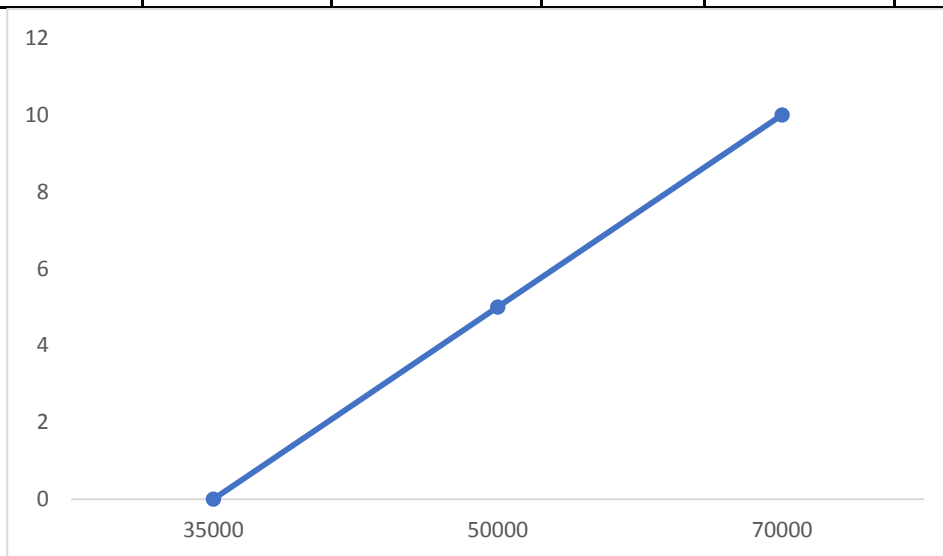


Рисунок 4.2 – Швидкодія програмного коду

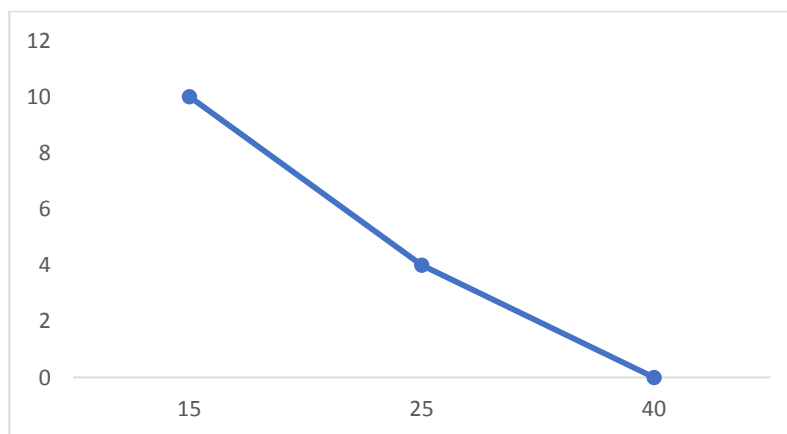


Рисунок 4.3 – Об'єм оперативної пам'яті

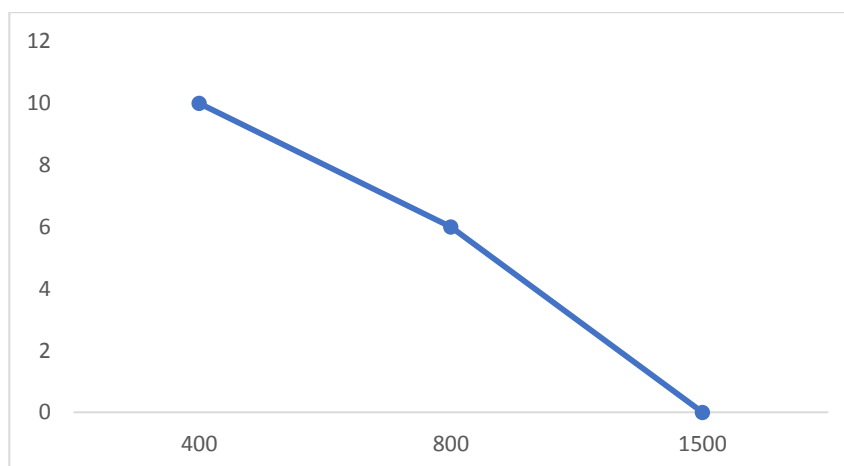


Рисунок 4.4 – Час обробки даних

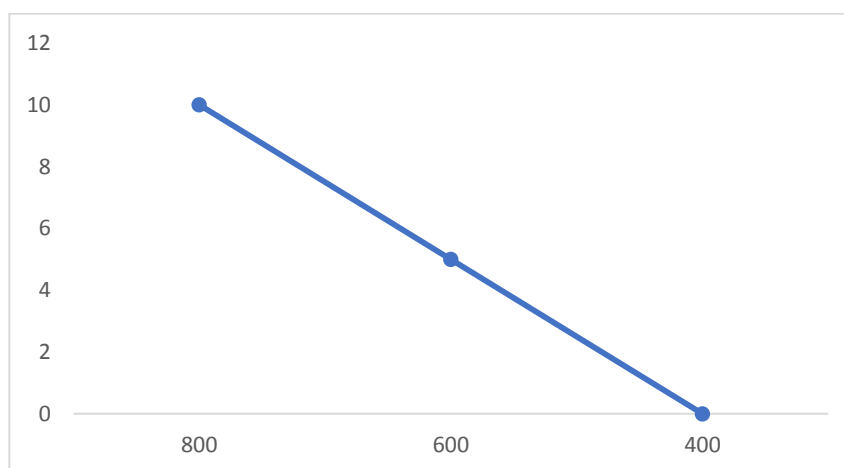


Рисунок 4.5–Потенційний об'єм коду

4.3.3 Експертне оцінювання параметрів

На цьому кроці кожному експерту необхідно оцінити ступінь важливості кожного параметру для поставленої цілі, в даному випадку це створення програмного забезпечення, яке буде найбільш адекватні моделі за встановленими критеріями, а також відповідають вимогам якості оцінки прогнозів.

Значення значимості для кожного параметру можна визначити за допомогою порівняння їх між собою, отримаємо порівняльну таблицю

Таблиця 4.3 – Результати ранжування параметрів

Позначення параметра	Назва параметра	Одиниці виміру	Ранг параметра за оцінкою експерта							Сума рангів R_i	Відхилення Δ_i	Δ_i^2
X1	Швидкість програмного коду	Оп/мс								12	5,5	0,25
X2	Об'єм оперативної пам'яті	Мб								23	5,5	0,25
X3	Час обробки даних	мс								26	8,5	2,25
X4	Потенційний об'єм коду	Кількість рядків коду								9	8,5	2,25
	Разом		0	0	0	0	0	0	0	70	0	0

Для розуміння того, що надані експертами оцінки, є достовірними, потрібно обчислити наступні показники:

а) сума рангів:

$$R_i = \sum_{j=1}^N r_{ij} R_{ij} = \frac{Nn(n+1)}{2} = 70, \quad (4.1)$$

б) середня сума рангів:

$$T = \frac{1}{n} R_{ij} = 17,5 \quad (4.2)$$

в) відхилення суми рангів по кожному параметру обчислюється за формулою:

$$\Delta_i = R_i - T \quad (4.3)$$

Для перевірки вірності обчислень використовується правило: сума усіх відхилень повинна бути рівні 0.

г) загальна сума квадратів відхилень:

$$S = \sum_{i=1}^N \Delta_i^2 = 205, \quad (4.4)$$

Далі необхідно обчислити коефіцієнт узгодженості, скористаємося такою формулою:

$$W = \frac{12S}{N^2(n^3-n)} = \frac{12 \cdot 205}{7^2(4^3-4)} = 0,84 \quad (4.5)$$

Отже, в нашому випадку коефіцієнт узгодженості дорівнює 0,84, що задовольняє вимогам до його значення (повинен бути більше, ніж 2/3).

Тепер попарно порівняємо наші параметри, зобразимо результати у вигляді таблиці:

Таблиця 4.4 – Порівняння параметрів

Параметри	Експерти							Кінцева оцінка	Числове значення
X1 і X2								<	0,5
X1 і X3								<	0,5
X1 і X4								>	1,5
X2 і X3								<	0,5
X2 і X4								>	1,5
X3 і X4								>	1,5

Визначимо числову інтерпретацію ступеню перевагу, для цього використаємо наступну формулу:

$$a_{ij} = \begin{cases} 1,5 & \text{при } X_i > X_j \\ 1,0 & \text{при } X_i = X_j \\ 0,5 & \text{при } X_i < X_j \end{cases} \quad (4.6)$$

Необхідно розрахувати ступінь вагомості для кожного з параметрів, для цього використаємо наступну формулу:

$$K_{Bi} = \frac{b_i}{\sum_{i=1}^n b_i},$$

(4.7)

де $b_i = \sum_{j=1}^N a_{ij}$; a_{ij} - ступінь переваги i -го параметра над j -тим.

Систематизуємо отримані значення у вигляді таблиці:

Таблиця 4.5 – Ступені вагомості параметрів

Параметр X_i	Параметри X_j				Перша ітер.		Друга ітер.		Третя ітер.	
	X1	X2	X3	X4	b_i	K_{Bi}	b_i^1	K_{Bi}^1	b_i^2	K_{Bi}^2
X1	,0	,5	,5	,5	4,5	0,219	2,625	0,216	1,10	0,215
X2	,5	,0	,5	,5	5,0	0,281	2,925	0,282	1,34	0,283
X3	,5	,5	,0	,5	6,0	0,344	3,725	0,347	1,59,5	0,348
X4	,5	,5	,5	,0	3,5	0,156	2,225	0,155	8,4,5	0,154
Всього:					19	1	115	1	488	1

4.3.4 Рівень якості альтернатив функцій

На цьому кроці необхідно розрахувати якість кожної альтернативи можливої реалізації функцій.

За наступною формулою визначимо коефіцієнт технічного рівня кожної альтернативи основних функцій:

$$K_K(j) = \sum_{i=1}^n K_{ei,j} B_{i,j}, \quad (4.9)$$

де n – кількість параметрів; K_{ei} – коефіцієнт вагомості i -го параметра; B_i – оцінка i -го параметра в балах.

Таблиця 4.6 – Рівень якості альтернатив основних функцій

Основні функції	Альтернатива	Абсолютне значення	Оцінка параметра	Коефіцієнт вагомості	Рівень якості
$F1(X1)$	A	60000	6,1	0,215	1,3115
$F2(X2)$	A	20	5,0	0,283	1,415
$F3(X3)$	B	500	8,3	0,348	2,8884
$F4(X4)$	A	700	2,6	0,154	0,4004
	B	450	8,7	0,154	1,3398

Отримані абсолютні значення за параметрами $X1$, $X2$ та $X3$ задовольняють встановленим технічним вимогам до програмного продукту, а отже є прийнятними.

Залишилося визначитись з вибором остаточно варіанту реалізації програмного продукту, для цього визначимо рівень якості для кожного з параметрів за наступною формулою:

$$K_K = K_{Ty}[F_{1k}] + K_{Ty}[F_{2k}] + \dots + K_{Ty}[F_{zk}] \quad (4.10)$$

Отримуємо наступні значення:

$$K_{K1} = 1,3115 + 1,415 + 2,8884 + 0,4004 = 6,0153$$

$$K_{K2} = 1,3115 + 1,415 + 2,8884 + 1,3398 = 6,9547$$

За розрахованими значеннями бачимо, що кращим є другий варіант, оскільки $6,9547 > 6,0153$, таким чином обираємо його.

4.4 Економічний аналіз варіантів розробки програмного продукту

Необхідно розрахувати працевтрати на розробку програмного продукту, цей процес включає такі завдання:

- Розробка проекту програмного продукту
- Розробка програмної оболонки

Завдання 1 за ступенем новизни відноситься до групи А, завдання 2 – до групи Б. За складністю алгоритми, які використовуються в завданні 1 належать до групи 1; а в завданні 2 – до групи 3.

Для реалізації завдання 1 використовується довідкова інформація, а завдання 2 використовує інформацію у вигляді даних.

Потрібно розрахувати витрати часу на виконання кожного із завдань. Загальні працевтрати розраховуємо за наступною формулою:

$$T_0 = T_P \cdot K_P \cdot K_{СК} \cdot K_M \cdot K_{СТ} \cdot K_{СТ.М}, \quad (4.11)$$

де T_P – трудомісткість розробки ПП;

K_P – поправочний коефіцієнт;

$K_{СК}$ – коефіцієнт на складність вхідної інформації;

K_M – коефіцієнт рівня мови програмування;

$K_{СТ}$ – коефіцієнт використання стандартних модулів і прикладних програм;

$K_{СТ.М}$ – коефіцієнт стандартного математичного забезпечення

Для першого завдання, виходячи із норм часу для завдань розрахункового характеру степеню новизни А та групи складності алгоритму 1, трудомісткість дорівнює: $T_p = 60$ людино-днів. Поправочний коефіцієнт, який враховує вид нормативно-довідкової інформації для першого завдання: $K_{П} = 1.7$. Поправочний коефіцієнт, який враховує складність контролю вхідної та вихідної інформації для всіх системи завдань рівний 1: $K_{СК} = 1$. Оскільки при розробці першого завдання використовуються стандартні модулі, врахуємо це за допомогою коефіцієнта $K_{СТ} = 0.8$. Тоді, за формулою 1.11, загальна трудомісткість програмування першого завдання дорівнює:

$$T_1 = 60 \cdot 1.7 \cdot 0.8 = 81,6 \text{ людино-днів.}$$

Для другого завдання (використовується алгоритм третьої групи складності, ступінь новизни Б), тобто $T_p = 15$ людино-днів, $K_{П} = 1.42$,

$K_{СК} = 1$, $K_{СТ} = 0.8$:

$$T_1 = 15 \cdot 1.42 \cdot 0.8 = 17,04 \text{ людино-днів.}$$

Складаємо трудомісткість відповідних завдань для кожного з обраних варіантів реалізації програми, щоб отримати їх трудомісткість:

$$T_I = (81,6 + 17,04 + 4,8 + 17,04) \cdot 8 = 963,84 \text{ людино-годин;}$$

$$T_{II}(81,6 + 17,04 + 6,91 + 17,04) \cdot 8 = 980,72 \text{ людино-годин};$$

Варіант II вимагає більших працевитрат.

В розробці беруть участь два програмісти з окладом 15000 грн., один аналітик з окладом 20000грн.Зарплата за годину визначається за наступною формулою::

$$C_{\text{ч}} = \frac{M}{T_m \cdot t} \quad (4.12)$$

де M – місячний оклад працівників;

T_m – кількість робочих днів у місяці;

t – кількість робочих годин в день.

$$C_{\text{ч}} = \frac{15000 + 15000 + 20000}{3 \cdot 21 \cdot 8} = 99,2 \text{ грн.}$$

Тоді, розрахуємо витрати на виплату заробітної плати, для цього скористаємося такою формулою:

$$C_{\text{ЗП}} = C_{\text{ч}} \cdot T_i \cdot K_d, \quad (4.13)$$

де $C_{\text{ч}}$ – величина погодинної оплати праці програміста;

T_i – трудомісткість відповідного завдання;

K_d – норматив, який враховує додаткову заробітну плату.

Зарплата розробників за варіантами становить:

$$\text{I. } C_{\text{ЗП}} = 99,2 \cdot 963,84 \cdot 1,2 = 114735,51 \text{ грн.}$$

$$\text{II. } C_{\text{ЗП}} = 99,2 \cdot 980,72 \cdot 1,2 = 116744,91 \text{ грн.}$$

Відрахування на єдиний соціальний внесок в залежності від групи професійного ризику (II клас) становить 22%:

$$\begin{aligned} \text{I. } C_{\text{ВІД}} &= C_{\text{ЗП}} \cdot 0.3677 = 114735,51 \cdot 0.22 = 25241,81 \text{ грн.} \\ \text{II. } C_{\text{ВІД}} &= C_{\text{ЗП}} \cdot 0.3677 = 116744,91 \cdot 0.22 = 25683,90 \text{ грн.} \end{aligned}$$

Тепер визначимо витрати на оплату однієї машино-години. (C_M)

Так як одна ЕОМ обслуговує одного програміста з окладом 15000 грн., з коефіцієнтом зайнятості 0,2 то для однієї машини отримаємо:

$$C_M = 12 \cdot M \cdot K_3 = 12 \cdot 15000 \cdot 0,2 = 36000 \text{ грн.}$$

З урахуванням додаткової заробітної плати:

$$C_{\text{ЗП}} = C_M \cdot (1 + K_3) = 36000 \cdot (1 + 0.2) = 43200 \text{ грн.}$$

Відрахування на єдиний соціальний внесок:

$$C_{\text{ВІД}} = C_{\text{ЗП}} \cdot 0.3677 = 43200 \cdot 0,22 = 9504 \text{ грн.}$$

Амортизаційні відрахування розраховуємо при амортизації 20% та вартості ЕОМ – 10000 грн.

$$C_A = K_{\text{ТМ}} \cdot K_A \cdot C_{\text{ПР}} = 1.15 \cdot 0.2 \cdot 10000 = 2300 \text{ грн.,}$$

де $K_{\text{ТМ}}$ – коефіцієнт витрат на транспортування та монтаж приладу;

K_A – річна норма амортизації;

$C_{\text{ПР}}$ – договірна ціна приладу.

Витрати на ремонт та профілактику розраховуємо як:

$$C_P = K_{\text{ТМ}} \cdot C_{\text{ПР}} \cdot K_P = 1.15 \cdot 10000 \cdot 0.05 = 575 \text{ грн.,}$$

де K_P – коефіцієнт витрат на поточні ремонти.

Ефективний годинний фонд часу ЕОМ за рік розраховуємо за формулою:

$$T_{\text{ЕФ}} = (D_K - D_B - D_C - D_P) \cdot t_3 \cdot K_B = (365 - 104 - 8 - 16) \cdot 8 \cdot 0.9 = 1706.4$$

годин,

де D_K – календарна кількість днів у році; D_B , D_C – відповідно кількість вихідних та святкових днів; D_P – кількість днів планових ремонтів

устаткування; t - кількість робочих годин в день; K_B - коефіцієнт використання приладу у часі протягом зміни.

Витрати на оплату електроенергії розраховуємо за формулою:

$$C_{\text{ЕЛ}} = T_{\text{ЕФ}} \cdot N_C \cdot K_3 \cdot C_{\text{ЕН}} = 1706,4 \cdot 0,32 \cdot 0,3 \cdot 0,27515 = 45,07 \text{ грн.},$$

де N_C – середньо-споживча потужність приладу;

K_3 – коефіцієнтом зайнятості приладу;

$C_{\text{ЕН}}$ – тариф за 1 КВт-годин електроенергії.

Накладні витрати розраховуємо за формулою:

$$C_H = C_{\text{ПР}} \cdot 0,67 = 10000 \cdot 0,67 = 6700 \text{ грн.}$$

Тоді, річні експлуатаційні витрати складатимуть:

$$C_{\text{ЕКС}} = C_{\text{ЗП}} + C_{\text{ВІД}} + C_A + C_P + C_{\text{ЕЛ}} + C_H \quad (4.14)$$

$$C_{\text{ЕКС}} = 43200 + 9504 + 2300 + 575 + 45,07 + 6700 = 62324,07 \text{ грн.}$$

Собівартість однієї машино-години ЕОМ дорівнюватиме:

$$C_{\text{М-Г}} = C_{\text{ЕКС}} / T_{\text{ЕФ}} = 62324,07 / 1706,4 = 36,52 \text{ грн/год.}$$

В рамках реалізації даного програмного продукту ведуться на ЕОМ, в такому випадку витрати на оплату машинного часу будуть розраховуватися за наступною формулою:

$$C_M = C_{\text{М-Г}} \cdot T \quad (4.15)$$

$$\text{I. } C_M = 36,52 \cdot 963,84 = 35199,44 \text{ грн.};$$

$$\text{II. } C_M = 36,52 \cdot 980,72 = 35815,89 \text{ грн.};$$

$$C_H = C_{\text{ЗП}} \cdot 0,67 \quad (4.16)$$

$$\text{I. } C_H = 114735,51 \cdot 0,67 = 76872,79 \text{ грн.};$$

$$\text{II. } C_H = 116744,91 \cdot 0,67 = 78219,09 \text{ грн.};$$

Таким чином, витрати на розробку програмного продукту будуть складати:

$$C_{\text{ПП}} = C_{\text{ЗП}} + C_{\text{ВІД}} + C_{\text{М}} + C_{\text{Н}} \quad (4.17)$$

$$\text{I. } C_{\text{ПП}} = 114735,51 + 25241,81 + 35199,44 + 76872,79 = 252049,55 \text{ грн.};$$

$$\text{II. } C_{\text{ПП}} = 116744,91 + 25683,90 + 35815,89 + 78219,09 = 256463,79 \text{ грн.};$$

4.5 Вибір кращого варіанта програмного продукту техніко-економічного рівня

Для визначення кращої альтернативи з техніко-економічної точки зору скористаємось наступною формулою:

$$K_{\text{ТЕР}j} = K_{\text{К}j} / C_{\text{Ф}j} \quad (4.18)$$

$$K_{\text{ТЕР}1} = 6,0153 / 252049,55 = 0,0000238$$

$$K_{\text{ТЕР}2} = 6,9547 / 256463,79 = 0,0000271;$$

Таким чином альтернатива 2 є більш ефективною з техніко-економічної точки зору.

Висновки до розділу

В цьому розділі було проведено функціонально-вартісний аналіз програмного продукту, що розробляється в рамках дипломного проектування. Процес аналізу складався з двох етапів.

На першому етапі проводився технічний аналіз програмного продукту, тобто були визначені основні функції продукту та альтернативні варіанти їх реалізації, а для визначення оптимальної альтернативи з технічної точки зору обчислювався рівень якості кожної з них.

На другому етапі виконувалося дослідження альтернатив на оптимальність з економічної точки зору, тут виконувалися обрахування витрат на трудові ресурси та матеріальне забезпечення, що необхідне для реалізації програмного продукту та визначався коефіцієнт ефективності для визначення кращої з альтернатив.

Таким чином, за результатом проведеного двоетапного аналізу, було визначено, що кращою є друга альтернатива, отже обрано такий варіант реалізації програмного продукту:

- Мова програмування – Python
- Зчитування даних з файлу
- Веб-інтерфейс
- Виведення результатів на екран з додатковою візуалізацією

ВИСНОВКИ ЗА РОБОТОЮ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Дипломна робота це аналіз та дослідження поведінки часових рядів, і в кінцевому вигляді – процес аналізу результатів побудованих регресійних та авторегресійних моделей.

У першій частині даної роботи було розглянуто та проаналізовано популярні явища у демографічній ситуації по всьому світу, зроблено огляд зміни основних демографічних характеристик, таких як: чисельність населення та темп його зростання, народжуваності та смертності. Оскільки однією з цілей дипломної роботи було в першу чергу, зануритись в першопричини демографічної ситуації України на сьогоднішній день, та зробити короткий огляд демографічної ситуації у світі. Дана дипломна робота була написана з метою аналізу короткострокових, середньострокових та довгострокових демографічних прогнозів народонаселення України та також з намаганням зобразити у найбільш наочними для сприйняття візуалізацій прогнозування демографічних процесів в Україні.

Під час виконання роботи було отримано декілька авторегресійних моделей та зроблено аналіз результатів цих побудованих моделей у графічному вигляді.

За результатами прогнозування чисельності населення України можливо зробити висновок, що не дуже добра динаміка щодо зменшення чисельності населення буде зберігатися і в майбутньому, а саме - чисельність населення і дедалі буде зменшуватись, також неможливо не зазначити великий по своєму впливу на цю ситуацію фактор міграції населення.

Зрозуміло що наявні всі признаки для необхідності підвищеної уваги до демографічної ситуації в Україні та втручання держави у вигляді вдосконалення демографічної політики.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Драган О.І. Демографія: [Електронний ресурс]: конспект лекцій для студентів напряму підготовки «Управління персоналом та економіка праці» денної форми навчання / О. І. Драган., Т.І. Тертична. – К.: НУХТ, 2014. – 132 с.
2. «Перспективи світового населення ООН» (2019). Department of Economic and Social Affairs, Population Division (2019). URL: <https://population.un.org/wpp/Download/Standard/Population/> (дата звернення: 12.05.2020).
3. Гладун О.М. Нариси з демографічної історії України ХХ століття : Монографія / О.М. Гладун ; НАН України, Ін-т демограф. та соціальн. дослідж. ім. М.В. Птухи.—Київ,2018. — 224 с.ISBN 978-966-02-8439-5
4. Комплексний демографічний прогноз України на період до 2050 р. (колектив авторів) / за ред. чл.-кор. НАНУ, д.е.н., проф. Е.М. Лібанової. – К.: Український центр соціальних реформ, 2006. – 138 с.
5. Позняк О. В. Демографічні перспективи України до 2060 року / О. В. Позняк, П. Є. Шевчук // Демографія та соціальна економіка. - 2014. - № 1. - С. 72-84. - Режим доступу: http://nbuv.gov.ua/UJRN/dse_2014_1_8.
6. Бідюк П. І. Ймовірісно-статистичні методи моделювання і прогнозування : [монографія] / П. І. Бідюк, О. П. Гожий. – Миколаїв : Чорноморський державний університет ім. Петра Могили, 2014. – 440 с.
7. Бідюк П.І., Романенко В.Д., Тимощук О.Л. Аналіз часових рядів. Київ: НТУУ «КПІ», Політехніка, 2013. 600 с.
8. Бідюк П.І., Баклан І.В., Литвиненко В.І. Моделювання і прогнозування гетероскедастичних процесів. Систем. дослідж. та інф. технології, 2004. С. 115-134.
9. Бідюк П.І., Баклан І.В., Рифа В.М. Системний підхід до побудови математичних моделей на основі часових рядів // Системні дослідження та інформаційні технології, № 3, 2002, с. 114-131.

10. Enders W. Applied econometric time series. - New York: John Wiley & Sons, Inc., 1995. - 434 p.
11. Дрейпер Н., Смит Г. Прикладной регрессионный анализ (т.2). - М.: Финансы и статистика, 1986. - 366 с.
12. Пашін В.П. Оцінка конкурентоспроможності електронних пристроїв на стадії проектування. - К. Економічний вісник НТУУ „КПІ”, 2006. - №3. с. 252-255.
13. Пашін В. П. Методичні вказівки до виконання економіко- організаційного розділу дипломних проектів (робіт) бакалаврів і спеціалістів для студентів Інституту прикладного системного аналізу: Навч. посібник / Пашін В. П., Романов В. В., Єгорова Н. В – К.: НТУУ “КПІ”, 2011. – 118 с.

ДОДАТОК А (ілюстративний матеріал доповіді)

Моделі демографічних процесів та прогнозування чисельності населення України

Дипломну роботу виконав :
Студент групи КА-63
Пихтін Олександр Костянтинович
Науковий керівник: В.о. завідувача
кафедри, доцент, Тимошук О.Л

Об'єкт, мета та предмет дослідження

- Демографічні процеси на території України
- Зробити аналіз наявних тенденцій у демографічній ситуації України, змодельовати прогноз ключових показників за 2020-2024 роки та оцінити його.
- Вибір структури підібраних мат. моделей, що з достатнім ступенем адекватності можуть описати демографічні процеси та гарантують оцінку прогнозу.

Задача

- Зробити збір статистичних даних, оцінити їх та попередньо обробити.
- Зробити аналіз часових рядів на наявність тренду.
- Зробити апроксимацію демографічних процесів використовуючи моделі авторегресії
- Зробити оцінку короткострокових прогнозів важливих демографічних показників (чисельність населення, народжуваність, смертність)
- Зробити аналіз отриманих прогнозованих значень для вищезазначених показників, та представити їх у вигляді графіків для наочності.

Чисельність населення планети

World and regional data	Total population in millions 2017
Arab States	359
Asia and the Pacific	3,960
Eastern Europe and Central Asia	243
Latin America and the Caribbean	641
East and Southern Africa	581
West and Central Africa	424
More developed regions	1,260
Less developed regions	6,290
Least developed countries	1,002
World	7,550

Застосовані критерії, за якими оцінювались побудовані моделі

- Для апроксимації демографічних процесів:
- Статистика DW (Дурбіна – Уотсона)
- SSE (Сума квадратів похибок)
- R^2 – коефіцієнт детермінації

$$DW = \frac{\sum_{t=2}^N (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^N \varepsilon_t^2}, \quad -SSE = \sum_{t=1}^N (y - \hat{y})^2$$

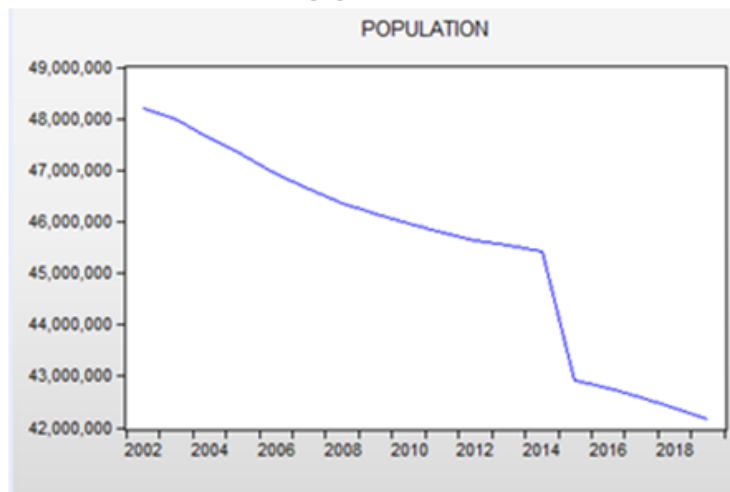
Застосовані критерії, за якими оцінювались побудовані моделі

Для прогнозування були використані:

- Середньоквадратична похибка прогнозу
- Середня абсолютна похибка прогнозу
- Коефіцієнт Тейла U

$$U = \frac{\sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - \hat{y}(k)]^2}}{\sqrt{\frac{1}{N} \sum_{k=1}^N y^2(k) + \frac{1}{N} \sum_{k=1}^N \hat{y}^2(k)}}$$

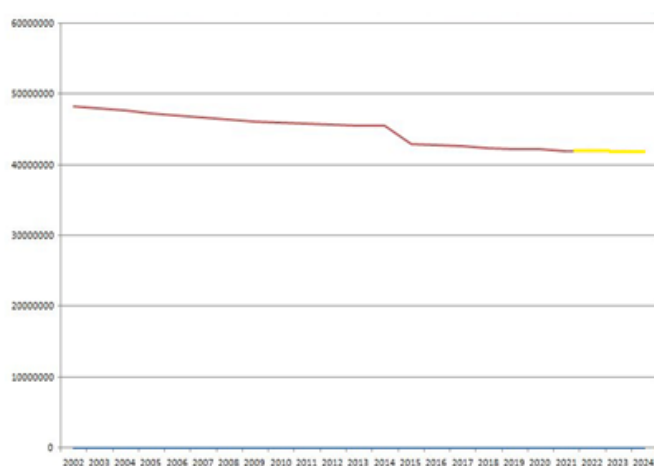
Стан чисельності населення України на 2002 - 2019 рр



Порівняння моделей для ряду чисельності населення

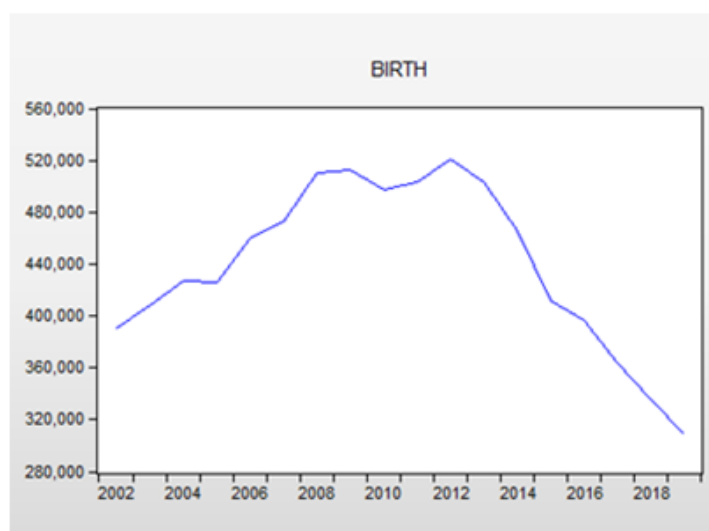
Модель	R^2	SSE	DW	Середн ьокв.ПЦ	Середн я АПП	U
AP(1)	0,861416	7,58E+12	1,542153	2485803	4.5668434	0.027824
AP(5)	0,892412	6,25E+12	1,832199	1434352	2.148206	0.025972
АРКС(6,2)	0,891145	6,17E+12	2,004136	1287793	2.108347	0.015365
АРКС(6,2)l	0.923456	-	2.003321	0.016545	0.090504	0.000894

Прогнозування з використанням моделі АРКС(6,2)



Рік	Чисельність населення
2020	42110413,34
2021	41836208,89
2022	41825318,02
2023	41761184,14
2024	41750323,78

Стан кількості народжених по роках в Україні на 2002 - 2019 рр



Народжуваність

Модель	R^2	SSE	DW	Середнь окв.ПП	Середн я АПП	U
AP(3)	0.8585468	5,91E+09	1.972342	45579.96	8.542422	0.048822
AP(7)	0.9192932	3,75E+09	1.883832	19405.28	3.393284	0.020443
АРКС(3,3)	0.9578123	1,46E+10	2.043421	25232.78	4.943232	0.028789
АРКС(7,7)	0.5638492	1.67E+10	0.308212	53231.22	11.74882	0.037372

Прогнозування з використанням моделі AP(7)

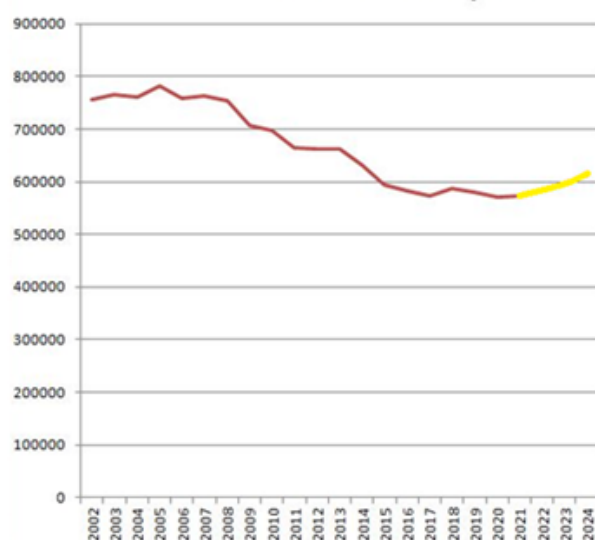


Рік	Прогнозо вані значення народжу ваності
2020	390210
2021	395201
2022	400245
2023	417829
2024	429452

Смертність

Модель	R^2	SSE	DW	Середн ьоквПП	Середн я АПП	U
AP(7)	0.954276	3.54E+09	2.021528	18486.45	2.411563	0.016123
АРКС(7,5)	0.988456	1.67E+09	1.784532	14116.54	1.820678	0.010915

Краща модель - АРКС(7.5)



Рік	Прогнозо вані значення народжує анності
2020	583152
2021	584669
2022	594218
2023	602161
2024	625214

Висновки

- Можна охарактеризувати демографічну ситуацію в Україні на 2020 – 2024 рр такими тезисами:
- Продожиться процес скорочення кількості населення
- Значення народжуванності по роках збільшиться
- Значення смертності по роках також збільшиться